

OM of: US-09-135-238B-2 to: EST:* out_format : pfs

Date: Sep 12, 2000 4:47 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters: -MODE=frame p2n.model -DEV=xlp

-Q/cgn2_1/USPTO_spo01/US09135238/runat_24082000_091113_16764/app_query.fasta_1.1282
-DB=EST -Qfmt=fastap -SURFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-FGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=4.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFWT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US09135238 -CGN1_1_2178 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-135-238B-2

Query length: 390

Database: EST:*

Database sequences: 5247842

Database length: -2090053206

Search time (sec): 1570.900000

score_list:

Sequence	Strd	Orig	25Score	EScore	Len	Documentation
gb_est25:AI760373	-	938.00	1321.43	8.9e-65	711	AI760373 wh87d11.xl NCI_CGAP_C
gb_est7:AA456707	+	544.00	778.77	1.5e-34	404	AA456707 aai3g04.rl Soares_NH
gb_est42:AA402953	+	551.00	771.54	3.8e-34	312	AA402953 UI-HF-BK0-abd-c-06-0-U
gb_est45:AW655210	+	522.00	735.08	4.1e-32	518	AW655210 105863 MARC_1BOV Bos
gb_est8:AA521993	+	497.00	697.66	5.0e-30	627	AA521993 vh61e12.rl Soares_mamm
gb_est45:AW655218	+	455.00	640.61	7.5e-27	494	AW655218 105873 MARC_1BOV Bos
gb_est42:AA444691	-	444.50	625.98	4.9e-26	482	AA444691 UI-H-B13-jdw-g-12-0-U
gb_est41:AW327067	-	415.00	587.24	7.0e-24	359	AW327067 20516 MARC_1BOV Bos
gb_est42:AA401870	+	372.00	524.91	2.1e-20	410	AA401870 UI-HF-BK0-aah-g-11-0-U
gb_est43:AA174968	+	360.00	506.70	2.1e-19	460	AA174968 ms78h03.rl Soares_mous
gb_est4:AA289667	+	349.00	492.13	1.4e-18	417	AA289667 vb20c07.rl Soares_mous
gb_est41:AA1508302	+	331.50	468.11	5.1e-17	568	AA1508302 vb34f04.yl Soares_mous
gb_est21:AA303944	-	305.00	428.19	5.1e-15	485	AA303944 oc58b10.sl NCI_CGAP_G
gb_est12:AA830123	+	299.00	430.06	4.0e-15	179	AA830123 aa46g04.rl NCI_CGAP_G
gb_est40:AA291041	-	292.50	408.96	6.0e-14	561	AA291041 UI-H-B12-agj-h-06-0-U
gb_est4:AA290194	+	251.00	351.96	8.9e-11	471	AA290194 vb34f04.rl Soares_mous
gb_est4:AA290194	+	251.00	351.96	8.9e-11	471	AA290194 vb34f04.rl Soares_mous
gb_est24:AA1743020	+	210.00	292.79	1.5e-07	521	AA1743020 wg8b10.xl Soares_NSF
gb_est8:AA509857	+	209.50	302.10	1.8e-07	521	AA509857 vh51a10.rl Soares_mamm
gb_est36:AW012833	+	180.00	255.12	2.2e-05	327	AW012833 uc03d11.yl Sugano_mous
gb_est16:AI137500	-	178.00	250.24	4.1e-05	398	AI137500 UI-R-C2p-oc-a-09-0-U
gb_est15:AI021484	-	157.50	227.35	0.0008	220	AI021484 ub08h03.rl Soares_mamm
gb_est43:AA177660	+	145.00	209.82	0.0074	216	AA177660 mt20d11.rl Soares_mous
gb_est49:AA840770	+	132.50	184.92	0.1800	431	H84070 ys94c09.rl Soares retina
gb_est46:AW731156	+	132.50	172.44	0.8922	1430	AW731156 GA_Ea0010C23 Gossyp
gb_est46:AW696459	-	128.00	172.11	0.9305	1203	AW696459 NF106D06ST1F1057 Deve
gb_est44:AW603666	-	128.00	174.86	0.6542	614	AW603666 CW0-CN0039-280100-168
gb_est26:AF14851	+	122.00	164.75	2.39	716	AF14851 wj68e01.xl NCI_CGAP_L
gb_est46:AF14851	+	121.00	172.18	0.9220	306	F14851 SSC24F11 Porcine small
gb_gst15:AZ033800	+	121.00	167.05	1.78	501	AZ033800 RPI-23-248K2-TV RPI-1
gb_est44:AW603303	+	120.50	164.23	2.56	614	AW603303 RCI-CN0012-170100-011
gb_est44:AW603303	+	120.50	161.83	3.48	722	AF074747 AF074747 DWR lab, Indi
gb_est14:AF074747	+	119.50	153.74	9.81	1467	AF074747 GA_Ea0010005 Gossyp
gb_est46:AW731219	-	119.00	156.46	6.93	1056	AF217085 AF217085 Oriza sativa
gb_est15:AF217085	+	116.50	151.76	12.66	1180	AF217085 GA_Ea0012K08 Gossyp
gb_est46:AW727502	+	116.50	151.41	13.23	1320	AW729443 GA_Ea0025A11 Gossyp
gb_est46:AW729443	+	115.00	156.60	6.80	604	AI260098 LP03783.5prime LP Dros
gb_est18:AI260098	-	115.00	148.47	19.29	1319	AW731173 GA_Ea0010G09 Gossyp
gb_est46:AW731173	+	113.00	153.71	9.85	607	AI257543 LP05944.5prime LP Dros
gb_est18:AI257543	-	113.00	153.31	10.37	631	AI260956 LP05008.5prime LP Dros
gb_est18:AI260956	-	113.00	153.29	10.39	632	AI260955 LP05007.5prime LP Dros
gb_est18:AI260955	-	113.00	153.29	10.39	631	AI260955 LP05007.5prime LP Dros

gb_est18:AI294256 - 113.00 153.29 10.39 632 ! AI294256 LP07584.5prime LP D
gb_est18:AI294053 - 113.00 153.26 10.44 634 ! AI294053 LP07346.5prime LP D
gb_est18:AI259828 - 113.00 153.05 10.72 647 ! AI259828 LP03332.5prime LP D
gb_est18:AI258031 - 113.00 152.97 10.83 652 ! AI258031 LP01114.5prime LP D

seq_name: gb_est25:AI760373

seq_documentation_block: 711 bp mRNA EST 20-DEC-1999
LOCUS AI760373
DEFINITION wh87d11.xl NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2387733 3'
similar to TR:060667 060667 ANTI-FAS-INDUCED APOPTOSIS. ; mRNA
sequence.
ACCESSION AI760373
VERSION AI760373.1 GI:5176040
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 711)
NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Oct 30, 1998 this sequence version replaced gi:3812483.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
M.D., Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1449 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 472.

FEATURES
source
1..711
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2387733"
/clone.lib="NCI_CGAP_CLL1"
/tissue_type="B-cell, chronic lymphocytic leukemia"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCATTCGCTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaudo."

BASE COUNT 127 a 195 c 239 g 148 t 2 others

ORIGIN

alignment_scores:

Quality: 938.00 Length: 202

Ratio: 4.860 Gaps: 1

Percent Similarity: 95.545 Percent Identity: 92.574

alignment_block:

US-09-135-238B-2 x AI760373/rev

Align seg 1/1 to reverse of: AI760373 from: 1 to: 711

190 ValSerArgAlaSerValAlaGlyAspLysProArgThrPheLeuPr 206

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

711 GTGTACAGCATCTCAAGTAGCAGGTGACAGCCGACCTCTTGCC 662

```

206 oSerThrThrAlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProG 223
|||||
661 ATCCACTACAGCTCAAAATAATCTCAGCTCTGGAGGGCTGCTCAAGCCCC 612
|||||
223 InThrPro. SerTyAsnHisHisThrArgLeuHisArgGlnArgAlaLe 239
|||||
611 AGACGCCAGGCTACACACACACACAGCGCTGCACAGCAGAGCACT 562
|||||
239 uAspTyrGlySerClnSerGlyArgGluGlnGlyPheHisIleLeuI 256
|||||
561 GAGCATATGCTTCAGCTCTGGAGGGAAGCCCAAGATTTCATCTCGA 512
|||||
256 leProThrIleLeuGlyLeuPheLeuAlaLeuLeuGlyLeuValVal 272
|||||
511 TCCCGACCATCTCTGGGCTTTTCCTGTCTGGCACTTCTGGGCTGGTG 462
|||||
273 LysArgAlaValGluArgArgLysAlaLeuSerArgAlaArgArgLe 289
|||||
461 AAAGGGCGCTTGAAGAGGAGAAAGCCCTCTCCAGCGGGCCCGCACT 412
|||||
289 uAlaValArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerP 306
|||||
411 GGCCTGTAGGATCGCGCCCTGGAGAGCTCCCAAGAGCCCGCGGGTGC 362
|||||
306 roArgProArgSerGlnAsnAsnIleTySerAlaCysProArgArgAla 322
|||||
361 CGCGACCGCGCTCCCAAAACACATCTACAGCGCTGCCCGCGCGGCT 312
|||||
323 ArgGlyAlaAspAlaGlyThrGlyAlaProValProGlyProG 339
|||||
311 CGTGGAGCGGCTGTGAGGACAGGAGGCGCCCGTTCCTGGCGCCGG 262
|||||
339 yAlaProLeuProAlaProLeuGlnValSerGluSerProTyrLeuH 356
|||||
261 AGCGCGCTGTGCCCGCCCGCTGCAGGTGTTGAATCTCCCTGGCTCC 212
|||||
356 iAlaProSerLeuLysThrSerCysGluTyrValSerLeuTyrHisGln 372
|||||
211 ATGCCCCATCTCTGAGACACAGCTGTGAATACGTGAGCTCTACACCA 162
|||||
373 ProAlaAlaMetMetGluAspSerAspSerAspTyrIleAsnValPr 389
|||||
161 CTTGCCGCATGATGGAGGACAGTATTCAGATGACTACATCAATGTTC 112
|||||
389 oAla 390
|||||
111 TGCC 108

seq_name: gb_est7:AA456707

seq_documentation_block:
LOCUS AA456707 404 bp mRNA EST 06-JUN-1997
DEFINITION aa13404.r1 Soares_NhMPu_S1 Homo sapiens cDNA clone IMAGE:813174
5', mRNA sequence.
ACCESSION AA456707
VERSION AA456707.1 GI:2179283
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
Moore,B., Scheinberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-Merck EST project 1997
JOURNAL Unpublished (1997)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1877745.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

```

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 390.

FEATURES

```

1..404
Location/Qualifiers
Source
1..404
/organism="Homo sapiens"
/db_xref="GB:6044084"
/db_xref="taxon:9606"
/clone="IMAGE:813174"
/clone_lib="Soares_NhMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Orgaan: mixed (see below); Vector: pMT3D-Pac
(Pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NbHPU, and fetal heart NbH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT 71 a 153 c 114 g 66 t
ORIGIN

```

```

alignment_scores:
Quality: 551.00 Length: 111
Ratio: 5.009 Gaps: 0
Percent Similarity: 99.099 Percent Identity: 98.198

```

alignment_block:

```

US-09-135-238B-2 x AA456707 ..
Align seg 1/1 to: AA456707 from: 1 to: 404

280 LysAlaLeuSerArgArgAlaArgLeuAlaValArgMetArgAlaLe 296
:::|||||
1 GAAGCCCTCTCCAGCGCGCGCGGCTGCGGTGAGGATGCGCGCCT 50
|||||
296 uGluSerSerGlnArgProArgGlySerProArgProArgSerGlnAsnA 313
|||||
51 GGAGAGCTCCAGAGCGCGCGGCTGCGCGGA. CGCGCTCCCAACA 99
|||||
313 snIleTyrSerAlaCysProArgArgAlaArgGlyAlaAspAlaGly 329
|||||
100 ACATCTACAGCGCTCCCGCGCGCGCTCGTGAGCGGAGCGTGCAGGC 149
|||||
330 ThrGlyGluAlaProValProGlyProGlyAlaProLeuProProAlaPr 346
|||||
150 ACAGGGGAGGCGCGCTTCCCGCGCGGAGCGCGTTCGCCCGCCGCC 199
|||||
346 oLeuGlnValSerGluSerProTyrLeuHisAlaProSerLeuLysThrS 363
|||||
200 GGTGAGGTGTCTGAATCTCCCTGGCTCCATGCCCTCTCTGAGACCA 249
|||||
363 ecCysGluTyrValSerLeuTyrHisGlnProAlaAlaMetMetGluAsp 379
|||||
250 GCTGTGAATACGTGAGCGCTTACCACCGAGCTGCGCGCATGATGGAG 299
|||||
380 SerAspSerAspTyrIleAsnValProAla 390
|||||
300 AGTGATTGAGTACTACATCAATGTTCTCTGCC 332
|||||
seq_name: gb_est42:AW402953
seq_documentation_block:

```

LOCUS AW402953 312 bp mRNA EST 16-FEB-2000
DEFINITION UI-HF-BKO-abd-c-06-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone IMAGE:3055859 5', mRNA sequence.
ACCESSION AW402953
VERSION AW402953.1 GI:6921743
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 312)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
On May 18, 1998 this sequence version replaced gi:3138550.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: www-bio.lnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.
Location/Qualifiers
1..312
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3055859"
/clone_lib="NIH_MGC_36"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LT1)"
/notes="Vector: pT73-Pac; Site_1: NotI; Site_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (0.5-1.5kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
97 a 74 c 87 g 54 t
BASE COUNT
ORIGIN
alignment_scores:
Quality: 544.00 Length: 101
Ratio: 5.386 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-135-238B-2 x AW402953 ..
Align seg 1/1 to: AW402953 from: 1 to: 312
41 GluMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyTh 57
10 GAATGTCATGTGAGGATATCTGTGCCGGGAGATGCTGGATCTGGAAC 59
57 rCysGlyThrValValSerThrThrAsnPhelLeLysAlaGluTyrLysG 74
60 ATGTGTTACCGTGGTATCCACCACCACTTCATCAAGCAGCAATACAGG 109
74 lyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 90
110 GCCGAGTTACTCTGAACAATAACCCACCAAGAATCTGTTCCCTAGTGGAG 159
91 ValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaG1 107
160 GTAACACAGCTGACAGAAGTGACAGCGGAGTCTATGCTCGGAGCGGG 209
107 yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 124

|||||
210 CATGACACAGACCGGGGAAGACCCAGAAAGTCACCCCTGAATGTCACA 259
124 erGluTyrGluProSerTrpGluGluGlnProMetProGluThrProLys 140
|||||
260 GTGAATACGAGCCATCATGGGAAGAGCAGCAATGCTGAGACTCCAAAA 309
141 Trp 141
||||
310 TGG 312
seq_name: gb_est45:AW655210
seq_documentation_block:
LOCUS AW655210 518 bp mRNA EST 05-APR-2000
DEFINITION 105863 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW655210
VERSION AW655210.1 GI:7421036
KEYWORDS EST.
SOURCE Bos taurus.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Bovidae; Bovinae; Bos.
1 (bases 1 to 518)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and Keele,J.W.
Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle
Unpublished (2000)
On Mar 8, 1999 this sequence version replaced gi:4388119.
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail@marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAG
Plate: 86 row: K column: 10
Seq primer: ATTAGGTGACATATAG.
Location/Qualifiers
1..518
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."
BASE COUNT 124 a 133 c 157 g 104 t
ORIGIN
alignment_scores:
Quality: 522.00 Length: 136
Ratio: 4.424 Gaps: 0
Percent Similarity: 86.765 Percent Identity: 68.382
alignment_block:
US-09-135-238B-2 x AW655210 ..
Align seg 1/1 to: AW655210 from: 1 to: 518
1 MetAspArgTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLe 17
|||||
110 ATGGACCTGTGGCTTTGGCGCTTACTTCTTCCTCCAGTAGTCGAGCCCC 159

```

17 uArglleuLeuProGluVallysValGluGlyGluLeuGlyGlySerValt 34
   ::::::::::::::::::::::::::::|
160 GAAGGTCTCCAGAGTAAGATGAAGGAATCTGGGAGGATCAATTA 209
   ::::::::::::::::::::::::::::::|
34 hrlleuGlyCysProLeuProGluMetHisValArglleuGlyLeuGlySerArg 50
   ::::::::::::::::::::::::::::::|
210 CCATCGAGGCCACTCTCCGAAACGATGTGAGGATATATCTGTGGCGG 259
   ::::::::::::::::::::::::::::::|
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67
   ::::::::::::::::::::::::::::::|
260 ACCATTGACAGAGTCTGGAAGATGCACCCGCTGGTGTCCAGCAACAATA 309
   ::::::::::::::::::::::::::::::|
67 eileLysAlaGluThrLysGlyArgValThrLeuLysGlnTyrProArgL 84
   ::::::::::::::::::::::::::::::|
310 CGTCAGAGAGGAATTCAGACCCGAGTCACGCTGGACGATGTCCGGACA 359
   ::::::::::::::::::::::::::::::|
84 ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
   ::::::::::::::::::::::::::::::|
360 GGAATCTGTCTTGTGTGATGACAGAGCTGACCAAGAATGACAGCGGG 409
   ::::::::::::::::::::::::::::::|
101 ValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117
   ::::::::::::::::::::::::::::::|
410 ATCTATGCTGTGGGGTGGCGGGAACACAGACGAGGAGCAAGCCAGCA 459
   ::::::::::::::::::::::::::::::|
117 sValThrLeuAsnValHisSerGluTyrGluProSerTyrGluGluGlnP 134
   ::::::::::::::::::::::::::::::|
460 GATCACCCCTGACTGTTCACAGCGTTTACGAGCCATCTCTGGGAAGAAG 509
   ::::::::::::::::::::::::::::::|
134 roMetPro 136
   ::::::::::|
510 CGATGCT 517
   ::::::::::|

```

seq_name: gb_est8:AA521993

```

seq_documentation_block: 627 bp mRNA EST 17-JUL-1997
LOCUS AA521993
DEFINITION v61e12.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
IMAGE:891502.5', mRNA sequence.
ACCESSION AA521993
VERSION AA521993.1 GI:2262738
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 627)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797443.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:519462
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 494.
Location/Qualifiers
1..627
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:891502"

```

FEATURES
source

```

/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Organ: mammary gland; Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site1: Not I;
Site2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5
TGTTACCAATCTGAAGTGGGAGGCGCGCAATGTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 165 a 192 c 156 g 114 t
ORIGIN
alignment_scores:
Quality: 497.00 Length: 206
Ratio: 3.227 Gaps: 5
Percent Similarity: 74.757 Percent Identity: 52.913
alignment_block:
US-09-135-238b-2 x AA521993 ..
Align seg 1/1 to: AA521993 from: 1 to: 627
104 CysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLe 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10 TGTGGTGTGGCATGAAGACACAGCAAGGCGACGCCAGAAATCACCT 59
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 uAsnValHisSerGluTyr...GluProSerTyrGluGlnProMetP 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 GAATCTCCATAATGAATACCCAGACCACTTCTGGGAAGATGAATGGACCT 109
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 roGluThrProLysTyrPheHisLeuProTyrLeuPheGlnMetPro... 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 CTGACGGCCCAAGATGGTTGCACAGATTCTGCAGCACCAAGATGCCCTGG 159
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 .....AlaTyrAlaSerSerSerLysPheValThrArgValTh 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 CTCCACGGGAGTGAACATCCAGCTCTCTGGAGTCATAGCCAAAGTAC 209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 rThrProAlaGlnArgGlyValProValHisHisSerSerProt 181
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 CACGCCAGCTCCAAAGACTGAGGCCCTCCGGTTCCACGAGCCTCCAGCA 259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 hrThrGlnIleThrHisArgProArgValSerArgAlaSerSerValala 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
260 TCACCTCAGTAACCCACATCCAGATTACAGAGCATTTCTGTGTCA 309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 GlyAspLysProArgThrPheLeuProSerThrThrAlaSerLysIleSe 214
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
310 GATACCAAGTCCCGAGCGCTCTGCCAGCAACCAAGCCTCAAGACTTC 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 rAlaLeuGluGlyLeuLeuLysProGlnThrPro.....SerTyrAsnH 229
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 CACTCAGCAGGCAAT.....CAGGCCCTTAGCGCAGCTACAGCC 399
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
229 iHisThrArgLeuHisArgGlnArgAlaLeuAspTyrGlySerGlnser 245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 ACCACACAGACTTCATGAGCAAGGACACGCCACCATGGCCCACTAT 449
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
246 GlyArgGluGlyGlnGly.....PheHisIleLe 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
450 GGGAGAGAGAGACCGAGGGCTTCACATCCCATCCCAAGATTTTCATCT 499
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
255 uIleProThrIleLeuGlyLeuPheLeuAlaLeuLeuGlyLeuValV 272
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
500 GATTCCGACCTTCTCGGCTTCTCTCTGTTCTTTGGGACTGGTGG 549
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


272 allysArgAlaValGluArgArgLysAlaLeuSerArgAlaArgArg 288
 |||||
 550 TAAAGAGCCATTCAAGGAGGAGAGCGCTCTCCAGACGTCGCGGCCGA 599
 |||||
 289 LeuAlaValArgMetArg 294
 |||||
 600 CTGGCGATGAGGAGCGCA 617

seq_name: gb_est45:AW655218

seq_documentation_block:
 LOCUS AW655218 494 bp mRNA EST 05-APR-2000
 DEFINITION 105873 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION AW655218
 VERSION AW655218.1 GI:7421044
 KEYWORDS EST.
 SOURCE Bos taurus.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 494)
 Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
 Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
 Keele,J.W.
 Design and use of four pooled tissue normalized cDNA libraries for
 EST discovery in cattle
 Unpublished (2000)
 On Mar 8, 1999 this sequence version replaced gi:4388128.
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@mail.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAACACCTATGACGAT
 BACKWARD: GTTTCCCGACGACGACG
 Plate: 86 row: M column: 8
 Seq primer: ATTAGTGACACTATAG.

FEATURES

source
 1..494
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 1BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from lymph node, ovary,
 fat, hypothalamus, and pituitary."
 BASE COUNT 118 a 125 c 150 g 101 t

alignment_scores:

Quality: 455.00 Length: 128
 Ratio: 4.213 Gaps: 0
 Percent Similarity: 84.375 Percent Identity: 64.844

alignment_block:

US-09-135-238B-2 x AW655218 ..

Align seg 1/1 to: AW655218 from: 1 to: 494

1 MetAspArgTrpLeuTrpProLeuTyrrPheLeuProValSerGlyAlaLe 17
 |||||
 111 ATGGACCTGTGGCTTTGGCGCTTTACTTCTCTCCAGTACTGCGGACCCC 160
 |||||
 17 uArgIleLeuProGluValGluGlyGluLeuGlySerValr 34

161 GAAGGTCCTTCCAGAAATAAGATGAAGAATGCTGGGAGGATCCATTA 210
 |||||
 34 hrIleLysCysProLeuProGluMetHisValArgIleTyrrLeuCysArg 50
 |||||
 211 CCATCGAGTCCCACTCTCTGAACGCATGTGAGGATATATCTGTGCCGG 260
 |||||
 51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67
 ::|||
 261 ACCATTGACGAGTCTGGAAGATGCACCCGCTGTCTCCAGCAACAATA 310
 |||||
 67 eIleLysAlaGluTyrrLysGlyArgValThrLeuLysGlnTyrrProArgL 84
 |||||
 311 CGTAAAGAGGAATCAAGCAGCGACGCTGGGACGATGTCCGGACA 360
 |||||
 84 yAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
 ::|||
 361 GGAATCTGTTCTTGGTGATGACAGAGCTGACCAAGAATGACAGCGGG 410
 |||||
 101 ValTyrrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnly 117
 |||||
 411 ATCTATGCTGTGGGTGGCGGGAACACAGACCGAGGCAAGACCATCA 460
 |||||
 117 sValThrLeuAsnValHisSerGluTyrrGluPro 128
 ::|||
 461 GATCACCTGACTGTTTACAGCGCTTACTAGCCA 494

seq_name: gb_est42:AW444691

seq_documentation_block:
 LOCUS AW444691 482 bp mRNA EST 17-FEB-2000
 DEFINITION UI-H-Bi3-ajw-g-12-0-UI.s1 NCI_CGAP_Sub5 Homo sapiens cDNA clone
 IMAGE:2733502 3', mRNA sequence.
 ACCESSION AW444691
 VERSION AW444691.1 GI:6986453
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 482)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
 NCI-CGAP clone distribution information can be found through the
 I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 The following repetitive
 elements were found in this cDNA sequence: 1-30,
 >POLY_A#Simple_repeat
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

source
 Location/Qualifiers
 1..482
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NCI_CGAP_Sub5"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub5
 is a subtracted library derived from NCI_CGAP_Sub4. The
 NCI_CGAP_Sub5 library had 3 million recombinants. A
 single-stranded DNA preparation of NCI_CGAP_Sub4 was used
 as a tracer in a subtractive hybridization with a driver
 comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM

3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)
1322376-1323911, 1456008-1456775, 1500552-1502855);
NCI CGAP Kid5 pool 1 LAM 3338-3342, 3722-3725, 3776-3778
(IMAGE Clones) 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI CGAP Lu5 pool 1 LAM 3575-3582,
3851-3854 (IMAGE Clones)
1414920-1417991, 1520904-1522439); NCI CGAP GC4 pool 1 LAM
3164-3167, 3716-3720, 3733-3735 (IMAGE Clones)
1257096-1258631, 1469064-1470983, 1475592-1476743);
NCI CGAP Pr22 pool 1 LAM 2457-2459, 2758-2759, 3062-3068
(IMAGE Clones) 985608-986759, 1101192-1101959,
1217928-1220615); NCI CGAP Co10 pool 1 LAM 2644-2653,
2871-2872 (IMAGE Clones)
1057416-1061255, 1144584-1145351). (10% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI CGAP Sub1 (IMAGE Clones) 2708616-2710535) and
NCI CGAP Sub2 (IMAGE Clones) 2710536-2712455) (10% of
the driver population), plus a pool of 11,136 clones from
NCI CGAP Sub3 (IMAGE Clones) 2712456-2723591) (10% of the
driver population), plus a pool of 5,472 clones from
NCI CGAP Sub4 (IMAGE Clones) 2723592-2728969) (70% of the
driver population). Subtraction was performed as
previously described [Bonaldi, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806. ;
TAG_LIB=NCI CGAP_Kid5; TAG_TISSUE=kidney; TAG_SEQ=ATTC";
BASE COUNT 103 a 107 c 111 g 161 t
ORIGIN

alignment_scores:
Quality: 444.50 Length: 117
Ratio: 4.401 Gaps: 3
Percent Similarity: 86.325 Percent Identity: 80.342

alignment_block:

US-09-135-238B-2 x AW444691/rev ..

Align seg 1/1 to reverse of: AW444691 from: 1 to: 482

41 GluMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyTh 57
|||||
481 GAAATGATGTGAGGATATCTGTCCGGGAGATGGCTGATCGGAAC 432
57 rCysGlyThrValSerThrThrAsnPhelIleLysAlaGluTyrLysG 74
|||||
431 ATGTGGTACCGTGGTATCCACCACCACTTCATCAAGCAGATACAGG 382
74 llyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 90
|||||
381 GCCGAGTTACTCTGAAGCAATACCCAGCAAGAATCTGCTCTAGTGGAG 332
91 ValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaG 107
331 GTAACACAGCTGACAGAAAGTGACAGCGGAGTCTATGCTCGGAGCGGG 282
107 yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 124
281 CATGACACAGACCGGGGAAAGACCCAGAAAGTCACCTGAATGTCACA 232
124 er...GluTyrGluProSerTrpGluGlnProMetProGluThrPro 139
231 GTGGTAGGTCTCCCGCTGATGGAGGCTCAG.....ACTACCCAG 191
140 LysTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSe 156
190 AAAATGTTTCACTTGGGAGATTG...GAACAGCCATCTTAACGTGAACAG 144
156 r 156
143 T 143
seq_name: gb_est41:AW327067

seq_documentation_block:
LOCUS AW327067 359 bp mRNA EST 27-JAN-2000
DEFINITION 20516 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW327067
VERSION AW327067.1 GI:6762988
SOURCE EST.
ORGANISM Bos taurus.
REFERENCE 1 (bases 1 to 359)
Smith, I.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M.,
Bennett, G.A., Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A. and
Keefe, J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT On Jan 6, 2000 this sequence version replaced gi:6676592.
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCGCCAGTCACGAGC
Plate: 13 row: K column: 5
Seq primer: ATTTAGGTGACACTATAG.
location/Qualifiers
1..359
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 90 a 111 c 87 g 71 t
ORIGIN

alignment_scores:
Quality: 415.00 Length: 119
Ratio: 4.192 Gaps: 1
Percent Similarity: 83.193 Percent Identity: 64.706
alignment_block:
US-09-135-238B-2 x AW327067 ..
Align seg 1/1 to: AW327067 from: 1 to: 359
76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92
|||||
2 GTCCAGCTGGACAGCTGTCGGACAGGAATCTGCTCTGTGTGTGATGAC 51
92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 109
|||||
52 AGAGCTGACCAAGAAATGACAGCGGAGTCTATGCTGTGGGGTGGCGGGA 101
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
|||||
102 ACACAGCCGAGGCAAGACCCAGACGATCACCTGACTGTCACAGCGTT 151
126 TyrGluProSerTrpGluGlnProMetProGluThrProLysTrpPh 142
|||||
152 TACGAGCCATCTCGGGAAGAGCGGATGCTGCTCCAGCTCCAGCATGGTT 201

```

142 e.....HisLeuProTyrLeuPheGlnMetProAlat 153
202 TATGATTCTTCAATGACATGCTCTGTTCCAGATGCCGTCAC 251
153 yfAlaSerSerSerLysPheValThrArgValThrThrProAlaGlnArg 169
252 ATGCCAGTCTTTAAATTCACGTCCAAAGTAACACACACAGCTCAAAG 301
170 GlyLysValProProValHisHisSerSerProThrThrGlnIleThrHi 186
302 ATCAAGTCTCTCCAGACACACAGGCTCCCAACCCCTCAGTCACCCA 351
186 sargPro 188
352 CGCGCCC 358

seq_name: gb_est42:AW401870

seq_documentation_block: 410 bp mRNA EST 16-FEB-2000
LOCUS AW401870
DEFINITION UT-HF-BK0-aah-g-11-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone
IMAGE:3053781 5', mRNA sequence.
ACCESSION AW401870
VERSION AW401870.1 GI:6920556
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 410)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT On Jun 22, 1998 this sequence version replaced gi:3246884.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.
FEATURES
source
1..410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3053781"
/clone_lib="NIH_MGC_36"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/lab_host="MGC85"
/notes="Vector: pT73-Pac; Site 1: NotI; Site 2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(0.5-1.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT 72 a 134 c 94 g 89 t 1 others
ORIGIN

alignment_scores:
Quality: 372.00 Length: 69
Ratio: 5.391 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-135-238B-2 x AW401870

```

```

Align seg 1/1 to: AW401870 from: 1 to: 410

322 AlaArgGlyAlaSerAlaGlyThrGlyGluAlaProValProGlyPr 338
|||||
10 GCTGCTGGAGCGGACGCTGCAGGACAGGAGCGCCCGTTCCTCCGCGCC 59
|||||
338 OGlyAlaProLeuProProAlaProLeuGlnValSerGluSerProTriPL 355
|||||
60 CGGAGCGCGGTTGCCCGCCCGCGCTGCAGGTCTCTGAATCTCCCTGGC 109
|||||
355 euHisAlaProSerLeuLysThrSerCysGluTyrValSerLeuTyrHis 371
|||||
110 TCCATGCCCATCTCTGAAGACCCAGCTGTGAATACGTGAGCCTTACCAC 159
|||||
372 GlnProAlaAlaMetMetGluAspSerAspSerAspTyrIleAsnVa 388
|||||
160 CAGCTGCCCGCATGATGGAGGACAGTATTCAGATGACTACATCATGT 209
|||||
388 lProAla 390
|||||
210 TCCCTGCC 216

seq_name: gb_est3:AA174968

seq_documentation_block: 460 bp mRNA EST 16-FEB-1997
LOCUS AA174968
DEFINITION ms78h03.r1 Soares mouse 3NBMS Mus musculus cDNA clone IMAGE:617717
IMAGE: 5', mRNA sequence.
ACCESSION AA174968
VERSION AA174968.1 GI:1756098
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 460)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HMNI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:378541
Seq primer: -28M13 rev2 from Amerisham
High quality sequence stop: 459.
FEATURES
source
1..460
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:617717"
/clone_lib="Soares mouse 3NBMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TCTTACCAATCTGAAGTGGAGCGCGCCGCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I

```

and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 108 a 158 c 117 g 77 t
ORIGIN

alignment_scores:
Quality: 360.00 Length: 153
Ratio: 3.130 Gaps: 5
Percent Similarity: 75.163 Percent Identity: 54.902

alignment_block:
US-09-135-238B-2 x AA174968 ..

Align seg 1/1 to: AA174968 from: 1 to: 460

```

188 ProArgValSerArgAlaSerSerValAlaGlyAspLysProArgThrPh 204
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
   7 CCCAGAGTTTACAGAGCATTTCTGTGTCAGCTACCAAGTCCCGCGCT 56

204 eLeuProSerThrThrAlaSerLysLysSerAlaLeuGlyLeuLeuL 221
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
   57 CTTGCCAGCAACACAGCCTCAGAGACTTCCACTCAGCAGCA...ATCA 103

221 ysProGlnThrProSerTyrAsnHisHisThrArgLeuHisArgGlnArg 237
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
   104 GGGCCCTAGAGCCAGCTACAGCCACACAGCAGCTTATGAGCAAGG 153

238 AlaLeuAspTyrGlySerGlnSerGlyArgGlyGlnGly..... 251
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
   154 ACAGCGCCAGCATGCGCCACACTATGGAGAGAGAGCGAGGCTTCACAT 203

252 .....PheHisLeuLeuLeuProThrIleLeuGlyLeuPheL 264
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
   204 CCCCATCCCAAGATTCACATCTGATCCGAGCTTCTGGGCTTCTCT 253

264 eLeuLeuLeuLeuValValLysArgAlaValAlaValGluArgLys 280
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
   254 TGCTGTGTTCTTGGGACTGTGTAAAGAGCCATCAAGAGGAGGA 303

281 AlaLeuSerArgAlaArgArgLeuAlaValArgMetArgAlaLeuG 297
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
   304 GCTTCTCCAGACGTGCGGGCGGACTAGCGATAGGAGCGAGCGGG 353

297 userSerGln.Arg.ProArgGlySerPro.....ArgProAr 309
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
   354 GGTTCCTCCCGCGTTCCTCCACAGCGCGGGGATGCTCGCAGCGCG 403

309 qSerGlnAsnAnilletYrSerAlaCyProArgArgAlaArgGlyAla 326
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
   404 CTCGACAGACACAGCTCTACAGCGCTGCCCGCGCGGCGGAGCCAG 453

326 spAla 327
   |||||
   454 ACAGC 458

```

seq_name: gb_est4:AA289667

seq_documentation_block: 417 bp mRNA EST 14-APR-1997
LOCUS AA289667
DEFINITION vb20c07.rl Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:749484
5', mRNA sequence.
ACCESSION AA289667
VERSION AA289667.1 GI:1937343
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 417)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterson,R.

TITLE
JOURNAL
COMMENT

The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:458468

Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 405.

FEATURES
source

1. 417
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:749484"
/clone_lib="Soares mouse 3NbMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGCTGTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."

BASE COUNT 109 a 140 c 91 g 77 t
ORIGIN

alignment_scores:
Quality: 349.00 Length: 138
Ratio: 3.231 Gaps: 2
Percent Similarity: 78.261 Percent Identity: 53.623

alignment_block:

US-09-135-238B-2 x AA289667 ..

Align seg 1/1 to: AA289667 from: 1 to: 417

```

162 ArgValThrThrProAlaGlnArgGlyLysValProValHisHis 178
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
   4 AAAGTTACCCAGCCCAACTCCAAGACTGAGGCCCTCCGGTTCACAGCC 53

178 rSerProThrThrGlnIleThrHisArgProArgValSerArgAlaSerS 195
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
   54 CTCAGCATCATCTTCAGTAACCCACATCCAGAGTTTACAGAGCATTT 103

195 erValAlaGlyAspLysProArgThrPheLeuProSerThrAlaSer 211
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
   104 CTGTGTACGTACCAAGTCCCGCGCTCTCGCAGCAACACAGCGCTCA 153

212 LysIleSerAlaLeuGluGlyLeuLysProGlnThrProSerTyrAs 228
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
   154 AAGCTTCATCTCAGCAGCA...ATCAGGCCCTAGAGCGCCAGCTACAG 200

228 nHisHisThrArgLeuHisArgGlnArgAlaLeuAspTyrGlySerGlnS 245
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
   201 CCACACACAGCAGCTTCATGAGCAAGGACACGCCACCATGGCCCACT 250

```

245 erGlyArgGluGlyGlnGly.....PheHisIle 254
 ||||| ||||| :|||
 251 ATGGGAGAGAACCGAGGCTTCACATCCCATCCAGAAATTCACATC 300
 ||||| ||||| :|||
 255 LeuIleProThrIleGlyLeuPheLeuLeuAlaLeuLeuGlyLeuVa 271
 ||||| ||||| :|||
 301 CTGATTCGACGCTTCCTGGGCTTCTCTGCTGGTCTTTGGGACTGGT 350
 ||||| ||||| :|||
 271 lVallysArgAlaValGluArgGlyAlaLeuSerArgAlaArgA 288
 ||||| ||||| :|||
 351 GGTAAGAAGACCATCAAGAGGAGAGGCTCTCCAGACGTCGGGCC 400
 ||||| ||||| :|||
 288 rgLeuAlaValarg 292
 ||||| ||||| :|||
 401 GACTAGCGATGAGG 414

seq_name: gb_est21:AI508502

seq_documentation_block:

LOCUS AI508502 568 bp mRNA EST 12-MAR-1999
 DEFINITION vb34f04.y1 Soares mouse lymph node NBLMLN Mus musculus cDNA clone
 IMAGE:750847 5', mRNA sequence.

ACCESSION AI508502

VERSION AI508502.1 GI:4407407

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 568)
 Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
 Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
 Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
 Ritter,E., Kohn,S., Shin,F., Jackson,F., Cardenas,M., McCann,R.,
 Waterston,R. and Wilson.R.
 The WashU-NCI Mouse EST Project 1999

Unpublished (1999)

On Nov 2, 1998 this sequence version replaced gi:3830418.

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:459831

This read is a RESEQUENCE of a previously sequenced mouse clone

This read has been verified (found to hit its original self in the

correct orientation)

Putative full length read

vector to vector length is 639

Seq primer: -40RP from Gibco

High quality sequence stop: 458.

FEATURES

source

1..568
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_image="IMAGE:750847"
 /clone_lib="Soares mouse lymph node NBLMLN"
 /sex="male"
 /tissue_type="lymph node"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Organ: Lymph node; Vector: pT7T3D-Pac (Pharmacia)
 with a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
 1st strand cDNA was primed with a Not I - oligo(dT) primer
 [5].
 TGTACCATCTGAGTGGGAGCGCGGCTACTTTTTTTTTTTTTTTTTTTTTT
 3]; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. RNA

provided by Dr. Bertrand Jordan. Library constructed and
 normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 109 a 191 c 144 g 124 t
 ORIGIN

alignment_scores:

Quality: 331.50 Length: 115

Ratio: 3.900 Gaps: 1

Percent Similarity: 73.913 Percent Identity: 60.000

alignment_block:

US-09-135-238B-2 x AI508502 ..

Align seg 1/1 to: AI508502 from: 1 to: 568

280 LysAlaLeuSerArgArgAlaArgLeuAlaValArgMetArgAlaLe 296

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

42 GAAGCTCTCTCCAGAGTGGGGCCGACTAGCGATGAGGAGCGAGCGCG 91

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

296 uGluSerSer.....GlnArgProArgGlySerProArgP 308

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

92 GGGGGCTTCCCGCCCGTTCCTCCACACAGCGCGGGATGCTCGCAGAGGC 141

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

308 rcArgSerGlnAsnAsnIleTyrSerAlaCysProArgArgAlaArgGly 324

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

142 CGCGCTCGCAGAACACACGCTCTACAGCGCTGCCCGCGCGCACGGGA 191

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

325 AlaAspAlaAlaGlyThrGlyGluAlaProValProGlyProGlyAlaPr 341

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

192 CCAGACAGCTGGTCCAGCGAGGCTCGCTCTCAGCCCGCCAGCCCTC 241

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

341 oLeuProAlaProLeuGlnValSerGluSerProTrpLeuHisAlaP 358

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

242 AGCGTCCCCGCTTCTCCGACGACTCTTGAAGCTCTCTGGCCCCACACC 291

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

358 roSerLeuLysThrSerCysGluTyrValSerLeuTyrHisGlnProAla 374

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

292 CATCTCTGAAGATGAGCTGTGAATACGTGAGCTTGGGCTACAGCCTGCT 341

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

375 AlaMetGluAlaSerSerAspTyrIleAsnValPro 389

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

342 GTCACCTGGAAGACCCCTGATTCAGATGATTCATATATTCCT 386

seq_name: gb_est12:AA830944

seq_documentation_block:

LOCUS AA830944 485 bp mRNA EST 07-APR-1998

DEFINITION oc58b10.sl NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1353883 3',

mRNA sequence.

ACCESSION AA830944

VERSION AA830944.1 GI:2904043

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 485)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,

Ph.D., Gerald Marti, M.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 983 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 473.
Location/Qualifiers

FEATURES

source
1. .485
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI); Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGGCGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 109 a 84 c 137 g 155 t

alignment_scores:

Quality: 305.00 Length: 59
Ratio: 5.169 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-135-238B-2 x AA830944/rev ..

Align seg 1/1 to reverse of: AA830944 from: 1 to: 485

179 SerProThrThrGlnIleThrHisArgProArgValSerArgAlaSerSe 195
|||||
485 TCCCCACACCAACCAATCAACCCCGCTCGAGTCCAGAGCATCTTC 436
|||||
195 rValAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerL 212
|||||
435 AGTAGCAGGTGTGAAGCCCGAACCTTCCTGCCATCCACTACAGCCTCAA 386
|||||
212 ySileSerAlaLeuGluGlyLeuLeuLysProGlnThrProSerTyrAsn 228
|||||
385 AAATCTCAGCTCTGGAGGGGTGCTCAAGCCCCCAGAGCCGCTACAC 336
|||||
229 HisHisThrArgLeuHisArgGlnArg 237
|||||
335 CACCACACAGGCTGCACAGGCAGG 309

seq_name: gb_est8:AA491123

seq_documentation_block:

LOCUS AA491123 179 bp mRNA EST 18-AUG-1997
DEFINITION aa46904.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:824022 5',
mRNA sequence.

ACCESSION AA491123

VERSION AA491123.1 GI:2220296

KEYWORDS EST.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 179)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncilogap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -28m13 rev1 ET from Amersham

High quality sequence stop: 116.

FEATURES

source

1. 179
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI); Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGGCGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 46 a 69 c 34 g 30 t

ORIGIN

alignment_scores:

Quality: 299.00 Length: 59
Ratio: 5.068 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-135-238B-2 x AA491123 ..

Align seg 1/1 to: AA491123 from: 1 to: 179

163 ValThrThrProAlaGlnArgGlyLysValProProValHisHisSerSe 179
|||||
3 GTTACACACAGCTCAAGGGGGAAGTCCCTCCAGTCCACCACTCTC 52
|||||
179 rProThrThrGlnIleThrHisArgProArgValSerArgAlaSerServ 196
|||||
53 CCCCACCAACCAATCAACCCCGCTCGAGTCTCCAGAGCATCTTCAG 102
|||||
196 aAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLys 212
|||||
103 TACAGGTGTGACAGCCCCGAGACCTTCCTGCCATCCACTACAGCCCAAAA 152
|||||
213 IleSerAlaLeuGluGlyLeuLeuLys 221
|||||
153 ATCTCAGCTCTGGAGGGGTGCTCAAG 179

seq_name: gb_est40:AW291041

seq_documentation_block:

LOCUS AW291041 561 bp mRNA EST 16-JAN-2000
DEFINITION UI-H-B12-adj-h-06-0-UI.s1 NCI_CGAP_Sub4 Homo sapiens cDNA clone
IMAGE:2724707 3', mRNA sequence.
ACCESSION AW291041
VERSION AW291041.1 GI:6697677

KEYWORDS SOURCE

EST.
human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

On May 18, 1998 this sequence version replaced gi:3137211.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. cDNA Library Preparation: M.B. Soares lab Clone distribution:

NCI-CGAP clone distribution information can be found through the

I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward

POLYA=Yes

FEATURES

source

Location/Qualifiers

1...561

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2724707"

/lab_host="NCI-CGAP-Sub4"

/note="vector: pT73D-Pac (Pharmacia) with a modified

polylinker. Site.1: Not I; Site.2: Eco RI; The

NCI-CGAP-Sub4 library is a subtracted library derived from

the NCI-CGAP-Sub2 library which is a subtracted library

derived from the NCI-CGAP-Sub1 library, which is a

subtracted library derived from BI. BI constitutes a

mixture of 21 normalized or subtracted NCI-CGAP

libraries: NCI-CGAP-Co4, NCI-CGAP-Pr22, NCI-CGAP-Pr28,

NCI-CGAP-Co10, NCI-CGAP-Co16, NCI-CGAP-Kid5,

NCI-CGAP-Kid12, NCI-CGAP-Kid3, NCI-CGAP-Kid11,

NCI-CGAP-Lym2, NCI-CGAP-Bt2, NCI-CGAP-Co8, NCI-CGAP-CLL1,

NCI-CGAP-Lei2, NCI-CGAP-Brn23, NCI-CGAP-Lu5,

NCI-CGAP-Lu24, NCI-CGAP-Lu19, NCI-CGAP-GC4, NCI-CGAP-GC6,

NCI-CGAP-Brn25. These 21 libraries were pooled and a

single-stranded DNA preparation of the resulting mixture

was used as a tracer in a subtractive hybridization with

a driver whose composition is detailed below:

NCI-CGAP-Kid3 pool 1 : LLAM 3334-3337, 3682-3683,

3798-3803 (IMAGE Clonoids 1322376-1323911,

1456008-1456775, 1500552-1502855) NCI-CGAP-Kid5 pool 1 :

LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids

1323912-1325831, 1471368-1472903, 1492104-1493255)

NCI-CGAP-Lu5 pool 1 : LLAM 3575-3582, 3851-3854 (IMAGE

Clonoids 1414920-1417991, 1520904-1522439) NCI-CGAP-GC4

pool 1 : LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE

Clonoids 1257096-1258631, 1469064-1470983,

1475592-1476743) NCI-CGAP-Pr22 pool 1 : LLAM 2457-2459,

2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759,

1101192-1101959, 1217928-1220615) NCI-CGAP-Co10 pool 1 :

LLAM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255,

1144584-1145351) Subtraction was performed as previously

described [Bonaldo, Lennon & Soares (1996): Normalization

and Subtraction: Two Approaches To Facilitate Gene

Discovery. Genome Research 6, 791-806.]

* TAG_LIB=NCI-CGAP-Co10; TAG_TISSUE=colon; TAG_SEQ=AAACG"

121 a 125 c 120 g 194 t 1 others

BASE COUNT

ORIGIN

alignment_scores:

Quality: 292.50

Ratio: 3.953

Percent Similarity: 81.319

Length: 91

Gaps: 3

Percent Identity: 72.527

alignment_block:

US-09-135-238B-2 x AW291041/rev ..

Align seg 1/1 to reverse of: AW291041 from: 1 to: 561

```

67 PheTleLysAlaGluTyrLysGlyArgValThrLeuLysGluTyrProAr 83
|||||
561 TTCATCAAGCGAGATCAAGGCCGAGTTACTCTGAAGCAATACCCAGC 512
|||||
83 gLyAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 100
|||||
511 CAAGAATCTCTTCTAGTGGAGGTAACACAGCTGACAGAAAGTGACAGCG 462
|||||
100 LyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 116
|||||
461 GAGTCTATGCTCGCGAGCGGGCATGACACAGACCGGNGAAGACCCAG 412
|||||
117 LysValThrLeuAsnValHisSer...GluTyrGluProSerTrpGluGl 132
|||||
411 AAAGTCAACCTGAATGTCCACAGTGGTAGTCTCCCTCGCTGATTGGAGGC 362
|||||
132 uGlnProMetProGluThrProLysTppPheHisLeuProTyrLeuPheG 149
|||||
361 TCAG.....ACTACCCAGAAATGTTTCACCTGGGAAGATTG...G 324
|||||
149 InMetProAlaTyrAlaSerSer 156
|||||
323 AACAGCCATCTAAGTGAACAGT 301

```


OM of: US-09-135-238B-2 to: GenEmbl.* out_format : pfs
Date: Sep 12, 2000 6:51 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+ p2n_model -DEV=xlp
-O/cgn2_1/USPTO.spool/US09135238/runat_24082000_091323_17880/app_query.fasta_1.1282
-DB=GenEmbl -OFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.500 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-XGAPOP=4.500 -XGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcet
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM_ext -MNLLEN=0 -MAXLEN=1000000
-USER=US09135238 @CGN1_1_5120 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-135-238B-2
Query length: 390
Database: GenEmbl.*
Database sequences: 972840
Database length: 892348106
Search time (sec): 2784.300000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_pr3:AF057557	+	2047.00	1733.42	1.9e+88	1339	AF057557 Homo sapiens anti-Fas
gb_htg11:AC023534	-	661.50	528.40	2.5e+21	221365	AC023534 Homo sapiens clone
gb_htg22:AC027119	-	284.00	220.68	0.0003	58628	AC027119 Homo sapiens chromos
gb_ro:XM006431	+	198.00	170.58	0.2098	3095	U06431 Mus musculus B6/CBAF1J
gb_cm:AF091137	+	192.00	166.16	0.3699	2848	AF091137 Trichosurus vulpecula
gb_ro:WMP1GR03	+	188.00	177.35	0.0880	417	U83428 Mus musculus polymeric I
gb_ro:WMP1GR3	+	188.00	175.90	0.1061	505	Y16525 Mus musculus pigR gene,
gb_pat:EL3253	+	188.00	145.10	5.51	29392	E15253 Mouse pig receptor gen
gb_ro:AB001489	+	188.00	145.10	5.51	29392	AB001489 Mus musculus DNA for
gb_pat:EL15470	+	187.50	165.39	0.4080	1911	E15470 cDNA encoding immunity
gb_ro:RNIGR	+	185.50	159.64	0.8532	3269	E15741 Rat mRNA for polymeric
gb_cm:AB032195	+	184.00	160.53	0.7613	2461	AB032195 Sus scrofa mRNA for p
gb_cm:BOVPTGR	+	183.00	156.96	1.20	3527	L04757 Bos taurus polymeric im
gb_pr2:S43435S03	+	182.00	173.73	0.1400	345	S43441 transmembrane secretory
gb_pat:A52094	+	182.00	160.66	0.7483	1936	A52094 Sequence 4 from Patent
gb_pat:A52093	+	182.00	160.30	0.7839	2031	A52093 Sequence 3 from Patent
gb_pat:A52091	+	182.00	159.02	0.9239	2405	A52091 Sequence 1 from Patent
gb_pr2:S62403	+	182.00	157.61	1.11	2898	S62403 transmembrane secretory
gb_pr1:HSRPR	+	182.00	157.55	1.12	2919	X73079 Homo sapiens encoding F
gb_cm:RFPGR1	+	182.00	155.90	1.38	3630	X81371 B.taurus mRNA for bovin
gb_ro:AF035963	+	171.00	149.35	3.19	2539	AF035963 Rattus norvegicus kid
gb_htg11:AC023534	-	167.00	112.12	378.14	221365	AC023534 Homo sapiens clone
gb_cm:OCIGRECR	+	166.50	143.09	7.13	3517	X00412 Rabbit mRNA for poly-im
gb_ba1:SCY7H	-	159.50	118.29	171.52	42655	AL109732 Streptomyces coelic
gb_pr1:HUMIGRPOLY	+	157.00	137.60	14.41	2526	M24559 Human poly-ig receptor
gb_ro:WMP1GR5	+	152.50	146.79	4.44	456	Y16527 Mus musculus pigR gene,
gb_pr2:HS238323	+	152.50	141.07	9.23	969	AJ238323 Homo sapiens mRNA for
gb_pr2:HS224864	+	152.50	141.07	9.24	970	AJ010101 Homo sapiens mRNA for
gb_pr2:HS224864	+	152.50	136.15	17.37	1857	AJ224864 Homo sapiens mRNA for
gb_pr2:HSU15177	+	147.50	108.02	640.04	43599	U15177 Human cosmid CRI-JC201
gb_htg2:ALL57395	+	145.00	94.11	3.8e+03	207067	AL157395 Homo sapiens chromo
gb_ro:WMP1GR05	+	143.00	138.38	13.03	481	U83430 Mus musculus polymeric I
gb_htg6:AC020151	-	142.50	102.95	1.2e+03	48836	AC020151 Drosophila melanog
gb_in2:AE003551	-	142.50	89.56	6.8e+03	285978	AE003551 Drosophila melanoga
gb_ba1:AB032367	-	142.00	100.37	1.7e+03	64957	AB032367 Streptomyces avermit
gb_ba2:AE001986	-	141.50	110.03	494.93	17183	AE001986 Deinococcus radiodur
gb_in1:LMFL2743	-	140.00	102.68	1.5e+03	38368	AL031910 Leishmania major Fri
gb_in1:SCN35	-	140.00	101.40	1.5e+03	45396	AL078610 Streptomyces coelic
gb_htg9:AC010208	+	140.00	91.53	5.3e+03	167015	AC010208 Homo sapiens clone
gb_htg26:AC009813	-	140.00	90.07	6.4e+03	202598	AC009813 Homo sapiens chromo

gb_ba2:AF047717 + 139.00 113.87 302.25 7836 ! AF047717 Streptomyces chrys
gb_ba1:SC9B5 - 138.50 101.82 1.4e+03 36388 ! AL035206 Streptomyces coel
gb_ov:GDCEBP + 137.50 124.25 79.84 1686 ! X66844 G.domesticus gene fo
gb_in2:AF227923 + 137.50 120.73 125.41 2683 ! AF227923 Tribolium castaneu
gb_ba1:DVAPSAB + 137.50 118.98 156.92 3379 ! Z69372 D.vulgaris apsa and
seq_name: gb_pr3:AF057557
seq_documentation_block:
LOCUS AF057557 1339 bp mRNA PRI 04-JUN-1998
DEFINITION Homo sapiens anti-Fas-induced apoptosis (TOSO) mRNA, complete cds.
ACCESSION AF057557
VERSION AF057557.1 GI:3169292
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1339)
AUTHORS Hitoshi,Y., Lorens,J., Kitada,S.I., Fisher,J., LaBarge,M.,
Ring,H.Z., Francke,U., Reed,J.C., Kinoshita,S. and Nolan,G.P.
TITLE Toso, a cell surface, specific regulator of Fas-induced apoptosis
in T cells
JOURNAL Immunity 8 (4), 461-471 (1998)
MEDLINE 98246048
REFERENCE 2 (bases 1 to 1339)
AUTHORS LaBarge,M. and Hitoshi,Y.
TITLE Direct Submission
JOURNAL Submitted (03-APR-1998) Molecular Pharmacology, Stanford
University, 300 Pasteur Dr., Palo Alto, CA 94304, USA
FEATURES
Source
1..1339
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q31-q32"
/cell_line="Jurkat"
/cell_type="T cell"
note="TOSO.2 cDNA; cloned by PCR using primers designed
from the original TOSO.1 cDNA sequence"
1..1339
/gene="TOSO"
20..1192
/gene="TOSO"
note="transmembrane protein; Ig superfamily member"
/codon_start=1
/product="anti-Fas-induced apoptosis"
/protein_id="AAC18830.1"
/db_xref="GI:3169293"
/translation="MDFWLWPLFLPVSGALRILPEVKVVEGLGGSVTKCPLPEMHV
RIYLCRMASGCGTVVSTNFIKAIEYKGRVTKQIPKKNLFLVEVTLQTESDGVY
ACAGMNTDRGKTKVHTLNSHSEYFSEWEPQMPETPKPHLPFLQMPAYASSKFV
TRVTPAQRGKVPVHVHSSPTQITHRPVSRASVAGDKPRTFLPSTAKISAEFG
LLKPQTPSNHHTRLHQRALDYGSQSGREGQGFHILPTILGLFLALLGLVVKRAV
ERRKALSRARLAVRMRALESSORPRGSPRPRSONNYSACPRRARGDAAGTGAP
VPGCAPLPAPLQVSESPWLHAPSLKTSCEYVSLRHQPMAMEDSDSDDYINVPA"
127
/gene="TOSO"
/note="occurs in TOSO.1 cDNA"
1015
/replace="a"
/gene="TOSO"
/note="occurs in TOSO.1 cDNA"
/replace="g"
BASE COUNT 309 a 432 c 340 g 258 t
ORIGIN
alignment_scores:
Quality: 2047.00 Length: 390
Ratio: 5.262 Gaps: 0
Percent Similarity: 99.744 Percent Identity: 99.744

alignment_block:

US-09-135-238B-2 x AF057557 ..

Align seg 1/1 to: AF057557 from: 1 to: 1339

```
1 MetAspArgTtpLeuTtpProLeuTyrPheLeuProValSerGlyAlaLe 17
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
20 ATGGAGCTTCCTGCTTTGGCCACTTTACTTCTCCAGTATCGGGGGCCCT 69
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
17 uArgfLeLeuProGluValLysValGluGlyLeuGlySerValT 34
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
70 GAGGATCCTCCAGAGTAAGGTAGAGGGGAGCTGGCGGATCAGTTA 119
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
34 hrLeuLysCysProLeuProGluMetHisValArgTleTyrLeuCysArg 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
120 CCATCAAGTGCCCACTTCTGAAATGCATGTGAGGATATATCTGTGGCGG 169
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
51 GluMetAlaCysGlyThrCysGlyThrValValSerThrThrAsnPh 67
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
170 GAGATGGCTGGATCTGGAACATGTGTACCGTGGTATCCACCACTT 219
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
67 eLeuLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAgl 84
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
220 CATCAAGCGCAATACAGAGCGCGAGTACTCTGAACCAATACCCACGCA 269
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
84 ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
270 AGAATCTGTCTTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGGA 319
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
101 ValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
320 GTCTATGCTCGGAGCGGCATGAACACAGACCGGGGAAGACCCAGAA 369
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
117 svalThrLeuAsnValHisSerGluTyrGluProSerTrrpGluGlnP 134
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
370 AGTCACCCCTGAATGTGCACAGTGAATACGAGCCATCATGGGAAGAGCAGC 419
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
134 roMetProGluThrProLysTrrpPheHisLeuProTyrLeuPheGlnMet 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
420 CAATGCTCGAGACTCAAAATGGTTTCATCTGCCCTATTTGTCAGATG 469
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
151 ProLatyAlaSerSerLysPheValThrArgValThrThrProAl 167
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
470 CCTGCATATGCCAGTTCTTCCAAATTCGTAACACAGAGTTTACCACACAGC 519
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
167 acLnaArgLysValProProValHisHisSerSerProThrThrGlnI 184
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
520 TCAAGGGGGCAAGTCCCTCCAGTTCACCACTCTCTCCCCACCCACCAAA 569
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
184 leThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLys 200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
570 TCACCCACCGCCCTCGAGTGTCCAGAGCATCTTCAGTAGCAGGTGACAG 619
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
201 ProArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuG1 217
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
620 CCCGCAACCTTCTCCCACTCCACTACAGCCTCAAAATCTCAGCTCTGGA 669
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
217 uGlyLeuLeuLysProGlnThrProSerTyrAsnHisThrArgLeuH 234
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
670 GGGGCTGCTCAAGCCCCAGAGCCGACGTACAAACACACACACACAGCTGC 719
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
234 lsArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGlyGln 250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
720 ACAGGCAGAGACACTGGACTATGGCTACAGTCTGGGAGGAAGGCCAA 769
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
251 GlyPheHisIleLeuIleProThrIleLeuGlyLeuPheLeuLeuAlaLe 267
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
770 GGAATTCACATCTCATCCGACATCTCTGGGCCCTTTTCTCTGCTGGCACT 819
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
267 uLeuGlyLeuValValLysArgAlaValGluArgArgLysAlaLeuSerA 284
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
820 TCTGGGCTGGTGGTGAAGAGGCCGTGTAAGAGGAGGAAGGCCCTCTCCA 869
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

```
284 rgArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
870 GCGGGCCCGCCGACTGGCGGTGAGGATGCGCGCCTGGAGAGTCCCGAG 919
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
301 ArgProArgGlySerProArgProArgSerGlnAsnAsnIleTyrSerAl 317
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
920 AGGCCCCCGGGTCCCGGACCGCGCTCCCAAAACAACATCTACAGCSC 969
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
317 acYrProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyClnAlap 334
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
970 CTGCCCCGCGCGCTCTGTGGAGCGAGCTGCAGGCACACAGGGGAAGCCC 1019
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
334 roValProGlyProGlyAlaProLeuProProAlaProLeuGlnValSer 350
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1020 CCGTTCGCCGCCCGCGAGCGCGTTCGCCCGCCCGCGCTCGAGGTGCT 1069
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
351 GluSerProTrrpLeuHisAlaProSerLeuLysThrSerCysGluTyrVa 367
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1070 GAATCTCCCTGGCTCCATGCCCCATCTCTGAGAGCCAGCTGTGAATACGT 1119
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
367 lSerLeuTyrHisGlnProAlaAlaMetMetGluAspSerAspSerAspA 384
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1120 GAGCCTCTACCAACCGCTGCGCCATGATGGAGGACAGTGTTCAGATG 1169
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
384 spTyrIleAsnValProAla 390
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1170 ACTACATCAATGTTCTCTGCC 1189
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
seq_name: gb_htg11:AC023534
seq_documentation_block:
LOCUS AC023534 221365 bp DNA HTG 02-MAR-2000
DEFINITION Homo sapiens clone RP11-462N18, *** SEQUENCING IN PROGRESS ***, 61
unordered pieces.
ACCESSION AC023534
VERSION AC023534.2 GI:7143452
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 221365)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome, clone RP11-462N18
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 221365)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castile,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeArelano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,Y., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivari,F.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
JOURNAL Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 2, 2000 this sequence version replaced gi:6978210.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
```

http://ftp.genome.washington.edu/RN/repeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center Project name: L5244
Center Clone name: 462_N_18

*** NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. the true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1108: contig of 1108 bp in length
* 1109 1208: gap of 100 bp
* 1209 2282: contig of 1084 bp in length
* 2293 2392: gap of 100 bp
* 2393 3673: contig of 1281 bp in length
* 3674 3773: gap of 100 bp
* 3774 4947: contig of 1174 bp in length
* 4948 5047: gap of 100 bp
* 5048 6104: contig of 1057 bp in length
* 6105 6204: gap of 100 bp
* 6205 7259: contig of 1055 bp in length
* 7260 7359: gap of 100 bp
* 7360 8667: contig of 1308 bp in length
* 8668 8767: gap of 100 bp
* 8768 10566: contig of 1799 bp in length
* 10567 10666: gap of 100 bp
* 10667 12173: contig of 1507 bp in length
* 12174 12273: gap of 100 bp
* 12274 13643: contig of 1370 bp in length
* 13644 13743: gap of 100 bp
* 13744 15215: contig of 1472 bp in length
* 15216 15315: gap of 100 bp
* 15316 16609: contig of 1294 bp in length
* 16610 16709: gap of 100 bp
* 16710 18085: contig of 1376 bp in length
* 18086 18185: gap of 100 bp
* 18186 19668: contig of 1483 bp in length
* 19669 19768: gap of 100 bp
* 19769 22083: contig of 2315 bp in length
* 22084 22183: gap of 100 bp
* 22184 24052: contig of 1869 bp in length
* 24053 24152: gap of 100 bp
* 24153 25902: contig of 1750 bp in length
* 25903 26002: gap of 100 bp
* 26003 27639: contig of 1637 bp in length
* 27640 27739: gap of 100 bp
* 27740 29377: contig of 1638 bp in length
* 29378 29477: gap of 100 bp
* 29478 30910: contig of 1433 bp in length
* 30911 31010: gap of 100 bp
* 31011 31225: contig of 215 bp in length
* 31226 31325: gap of 100 bp
* 31326 32890: contig of 1565 bp in length
* 32891 32990: gap of 100 bp
* 32991 34812: contig of 1822 bp in length
* 34813 34912: gap of 100 bp
* 34913 36817: contig of 1905 bp in length
* 36818 36917: gap of 100 bp
* 36918 38575: contig of 1658 bp in length
* 38576 38675: gap of 100 bp
* 38676 41309: contig of 2634 bp in length
* 41310 41409: gap of 100 bp
* 41410 43344: contig of 1935 bp in length
* 43345 43444: gap of 100 bp
* 43445 45853: contig of 2409 bp in length

* 45854 45953: gap of 100 bp
* 45954 48434: contig of 2481 bp in length
* 48435 48534: gap of 100 bp
* 48535 51026: contig of 2492 bp in length
* 51027 51126: gap of 100 bp
* 51127 54369: contig of 3243 bp in length
* 54370 54469: gap of 100 bp
* 54470 56661: contig of 2192 bp in length
* 56662 56761: gap of 100 bp
* 56762 59097: contig of 2336 bp in length
* 59098 59197: gap of 100 bp
* 59198 61551: contig of 2354 bp in length
* 61552 61651: gap of 100 bp
* 61652 64257: contig of 2606 bp in length
* 64258 64357: gap of 100 bp
* 64358 67511: contig of 3154 bp in length
* 67512 67611: gap of 100 bp
* 67612 70329: contig of 2718 bp in length
* 70330 70429: gap of 100 bp
* 70430 73572: contig of 3143 bp in length
* 73573 73672: gap of 100 bp
* 73673 77203: contig of 3531 bp in length
* 77204 77303: gap of 100 bp
* 77304 80599: contig of 3296 bp in length
* 80600 80699: gap of 100 bp
* 80700 84046: contig of 3347 bp in length
* 84047 84146: gap of 100 bp
* 84147 88357: contig of 4211 bp in length
* 88358 88457: gap of 100 bp
* 88458 92454: contig of 3997 bp in length
* 92455 92554: gap of 100 bp
* 92555 96892: contig of 4338 bp in length
* 96893 96992: gap of 100 bp
* 96993 101334: contig of 4342 bp in length
* 101335 101434: gap of 100 bp
* 101435 105438: contig of 4004 bp in length
* 105439 105538: gap of 100 bp
* 105539 110413: contig of 4875 bp in length
* 110414 110513: gap of 100 bp
* 110514 115994: contig of 5481 bp in length
* 115995 116094: gap of 100 bp
* 116095 120447: contig of 4353 bp in length
* 120448 120547: gap of 100 bp
* 120548 124483: contig of 3936 bp in length
* 124484 124583: gap of 100 bp
* 124584 130249: contig of 5666 bp in length
* 130250 130349: gap of 100 bp
* 130350 135462: contig of 5113 bp in length
* 135463 135562: gap of 100 bp
* 135563 139847: contig of 4285 bp in length
* 139848 139947: gap of 100 bp
* 139948 146731: contig of 6784 bp in length
* 146732 146831: gap of 100 bp
* 146832 151469: contig of 4638 bp in length
* 151470 151569: gap of 100 bp
* 151570 159901: contig of 8332 bp in length
* 159902 160001: gap of 100 bp
* 160002 168308: contig of 8307 bp in length
* 168309 168408: gap of 100 bp
* 168409 176195: contig of 7787 bp in length
* 176196 176295: gap of 100 bp
* 176296 189303: contig of 13008 bp in length
* 189304 189403: gap of 100 bp
* 189404 202135: contig of 12732 bp in length
* 202136 202235: gap of 100 bp
* 202236 221365: contig of 19130 bp in length.

FEATURES

source

1. .221365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-462N18"
misc_feature 1. .1108
Location/Qualifiers

```

misc_feature /note="assembly_fragment"
1209..2292
/note="assembly_fragment"
2393..3673
misc_feature /note="assembly_fragment"
3774..4947
/note="assembly_fragment"
5048..6104
misc_feature /note="assembly_fragment"
6205..7259
/note="assembly_fragment"
7360..8667

alignment_scores:
  Quality: 661.50      Length: 508
  Ratio: 3.410         Gaps: 8
Percent Similarity: 38.189      Percent Identity: 34.252

alignment_block:
US-09-135-238B-2 x AC023534 ..
Align seg 1/1 to: AC023534 from: 1 to: 221365

2 AspArgTrpLeuTrp..... 6
33094 GACGGCTGGAGCTGGGAAGTAGATGGGCTGGGATATAGTTAGAGATGGG 33143
|||||
7 ....ProLeuTyrPheLeuPro.....ValSerGlyAlaLeuArgI 19
|||||
33144 GTCCTCTGTCAGTAATCTCTCTCTCCCTGTCAGTATGGGGCCCTGAGGA 33193
|||||
19 leLeuProGluValIcLuGlyLeuLeuGlyGlySerValThrIle 35
|||||
33194 TCCTCCAGAAAGTAGAGGGGAGCTGGCGGATCAGTTACCATC 33243
|||||
36 LysCysProLeuProGluMetHisValArgIleTyrLeuCysArgGluMe 52
|||||
33244 AAGTGCCCACTTCCTGAATGTCATGTGAGGATATCTGTGCGGGAGAT 33293
|||||
52 tAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPheIleL 69
|||||
33294 GGCTGGATCTGGAACATGGTACCGTGTATCCACCACTTCATCA 33343
|||||
69 ysAlaGluTyrIcLysGlyArgValThrLeuLysGlnTyrProArgLysAsn 85
|||||
33344 AGCAGAAATCAGGGCCGAGTTACTCTGAAGCAATACCCACGCAAGAT 33393
|||||
86 LeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTy 102
|||||
33394 CTGTTCTAGTGGAGTAAACACAGCTGACAGAAAGTGACAGGGAGTCTA 33443
|||||
102 rAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValT 119
|||||
33444 TGCTCGGAGGGGGCATGACACAGACCGGGGAAGACCCAGAAAGTCA 33493
|||||
119 hrLeuAsnValHis..... 123
|||||
33494 CCCTGAATGTGCACAGTGGTAGGTTCTCTCCGCTGANTGGAGCTCAGACT 33543
|||||
123 ..... 123
33544 ACCAGAAATGTTTCATCTGGGAGATAGGAACACACATCTAATCTGTA 33593
|||||
123 ..... 123
33594 CAGAGTCTTAGACTCTTACACTGTGGCCATATAGCCACATATATGATTA 33643
|||||
123 ..... 123
33644 GATCCATATGAAACCTCATAGTCAATTACCTCCATGTGTCACACAAATTT 33693
|||||
123 ..... 123

33694 CTACCCACATGTATGAAAGAACAGATAGGTGCTCTCTTGTATGTTT 33743
123 ..... 123
33744 CATCTTACCAGGCATGAGGGAGTAGAGGCTTGAGATTGAGAGAGGC 33793
123 ..... 123
33794 TCTGAGATGGAGAAATTGGAAGCAAAAGAAATAACTATCAACGTAAT 33843
123 ..... 123
33844 TGTCCCATGAGAAATAAGAGAAAGACACTAATTTATCCAGCCACATTTAT 33893
123 ..... 123
33894 AGGCCAATCTTATTACTCTGTGACCTGATCTTTATTCTCTGGAACCC 33943
123 ..... 123
33944 GCCTTTCTCTCTCTCTCCACTGAGTTCCTCCACACATATGCACAGTAAT 33993
123 ..... 123
33994 CAACAGCCTAGAAATCCCAATGACCTCTGCTGTTTTCAGGGAATGTTGA 34043
123 ..... 123
34044 CTGACTAGGGGAACCAAGTATTTCTCCAGGGAAGCCCAACCCAGCATC 34093
123 ..... 123
34094 CTCTCTGCTCTTTTCTGAGTGGACCTGCTCCCTCATCCCTCACTGTCATCC 34143
123 ..... 123
34144 ACTTTGGCTGAATAGCTTCAGTTTACACCCAGCCCTCAGTCTCCCCCTA 34193
124 .....Ser..GluTyrGluProSerTr 130
|||||
34194 GAGACCAACCAATATTCCTCTCTCTTCCATAGATAAGGACCATCATG 34243
|||||
130 pGluGluGlnProMetProGluThrProLysTrpPheHisLeuProTyrIL 147
|||||
34244 GGAAGACAGCAATGCTGAGACTCCAAATGGTTTCACTGCCCCATT 34293
|||||
147 euPheGlnMetProAlaTyrAlaSerSerLysPheValThrArg... 162
|||||
34294 TGTTCAGATGCTGCATATGCCAGTTCTTCCAAATTCGTAACCAAGGT 34343
|||||
163 ValThrThrProAlaGlnArgGlyLysValPro..... 173
|||||
34344 CAGTTCCCATAGCAAGGGTAGGAGAGTGGGTGAGATGCTGTGGCAATG 34393
|||||
174 .....ProValHisHisSerSerProThrT 182
|||||
34394 TGAATATGAGGAGGGGGTCTATGTTCAGTTTCAT..... 34427
182 hrGlnIleThrHisArgProArgValSerArgAlaSerSerValAlaGly 198
|||||
34428 ..AGATGGCTGGCAGAGCTGAGTCTTAGTGTGTATTCCTGCC 34475
|||||
199 AspLysProArgThrPheLeuProSerThrThrAlaSerLysIleSerAl 215
|||||
34476 TTTAGAGGGCACAATACCTCCCTGAGATCTGGGGTCCCATATCCTTCAAG 34525
|||||
215 aLeuGluGlyLeuLeuLysProGlnThrProSerTyrAsnHisHis.... 230
|||||
34526 CCTGTCCGACAGCTCAGGTCCCTGGTCTCTAGTGCAGAACTTACATATTA 34575
|||||
231 ..ThrArgLeuHisArgGlnArg 237
|||||
34576 CCACAGGTGGCACCACCAAGCAA 34598

```

```

seq_name: gb_htg22:AC027719
seq_documentation_block:
LOCUS      AC027719      58628 bp      DNA      HTG      01-APR-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-258P9 map 8, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION  AC027719
VERSION    AC027719.1 GI:7382531
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 58628)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castile,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Melarim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zalnouh,J., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence.submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L7449
            Center clone name: 258_P_9
            -----
            * NOTE: This record contains 74 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *
            * 1 671: contig of 671 bp in length
            * 672 771: gap of 100 bp
            * 772 1469: contig of 698 bp in length
            * 1470 1569: gap of 100 bp
            * 1570 2261: contig of 692 bp in length
            * 2262 2361: gap of 100 bp
            * 2362 3051: contig of 690 bp in length
            *
            * 3052 3151: gap of 100 bp
            * 3152 3816: contig of 665 bp in length
            * 3817 3916: gap of 100 bp
            * 3917 4597: contig of 681 bp in length
            * 4598 4697: gap of 100 bp
            * 4698 5395: contig of 698 bp in length
            * 5396 5495: gap of 100 bp
            * 5496 6185: contig of 690 bp in length
            * 6186 6285: gap of 100 bp
            * 6286 6969: contig of 684 bp in length
            * 6970 7069: gap of 100 bp
            * 7070 7780: contig of 711 bp in length
            * 7781 8590: contig of 710 bp in length
            * 8591 8690: gap of 100 bp
            * 8691 9372: contig of 682 bp in length
            * 9373 9472: gap of 100 bp
            * 9473 10162: contig of 690 bp in length
            * 10163 10262: gap of 100 bp
            * 10263 10944: contig of 682 bp in length
            * 10945 11044: gap of 100 bp
            * 11045 11747: contig of 703 bp in length
            * 11748 11847: gap of 100 bp
            * 11848 12547: contig of 700 bp in length
            * 12548 12647: gap of 100 bp
            * 12648 13336: contig of 689 bp in length
            * 13337 13436: gap of 100 bp
            * 13437 14103: contig of 667 bp in length
            * 14104 14203: gap of 100 bp
            * 14204 14872: contig of 669 bp in length
            * 14873 14972: gap of 100 bp
            * 14973 15682: contig of 710 bp in length
            * 15683 15782: gap of 100 bp
            * 15783 16491: contig of 709 bp in length
            * 16492 16591: gap of 100 bp
            * 16592 17290: contig of 699 bp in length
            * 17291 17390: gap of 100 bp
            * 17391 18088: contig of 698 bp in length
            * 18089 18188: gap of 100 bp
            * 18189 18883: contig of 695 bp in length
            * 18884 18983: gap of 100 bp
            * 18984 19671: contig of 688 bp in length
            * 19672 19771: gap of 100 bp
            * 19772 20413: contig of 642 bp in length
            * 20414 20513: gap of 100 bp
            * 20514 21196: contig of 683 bp in length
            * 21197 21296: gap of 100 bp
            * 21297 21992: contig of 696 bp in length
            * 21993 22092: gap of 100 bp
            * 22093 22804: contig of 712 bp in length
            * 22805 22904: gap of 100 bp
            * 22905 23604: contig of 700 bp in length
            * 23605 23704: gap of 100 bp
            * 23705 24402: contig of 698 bp in length
            * 24403 24502: gap of 100 bp
            * 24503 25191: contig of 689 bp in length
            * 25192 25291: gap of 100 bp
            * 25292 25974: contig of 683 bp in length
            * 25975 26074: gap of 100 bp
            * 26075 26765: contig of 695 bp in length
            * 26770 26869: gap of 100 bp
            * 26870 27572: contig of 703 bp in length
            * 27573 27672: gap of 100 bp
            * 27673 28373: contig of 701 bp in length
            * 28374 28473: gap of 100 bp
            * 28474 29165: contig of 692 bp in length
            * 29166 29265: gap of 100 bp
            * 29266 29956: contig of 691 bp in length
            * 29957 30056: gap of 100 bp
            * 30057 30741: contig of 685 bp in length
            * 30742 30841: gap of 100 bp
            * 30842 31513: contig of 678 bp in length
            * 31520 31619: gap of 100 bp

```

31620 32323: contig of 704 bp in length
32324 32423: gap of 100 bp
32424 33110: contig of 687 bp in length
33111 33210: gap of 100 bp
33211 33902: contig of 692 bp in length
33903 34002: gap of 100 bp
34003 34708: contig of 706 bp in length
34709 34808: gap of 100 bp
34809 35500: contig of 692 bp in length
35501 35600: gap of 100 bp
35601 36300: contig of 700 bp in length
36301 36400: gap of 100 bp
36401 37097: contig of 697 bp in length
37098 37197: gap of 100 bp
37198 37900: contig of 703 bp in length
37901 38000: gap of 100 bp
38001 38695: contig of 695 bp in length
38696 38795: gap of 100 bp
38796 39487: contig of 692 bp in length
39488 39587: gap of 100 bp
39588 40291: contig of 704 bp in length
40292 40391: gap of 100 bp
40392 41101: contig of 710 bp in length
41102 41201: gap of 100 bp
41202 41886: contig of 685 bp in length
41887 41986: gap of 100 bp
41987 42686: contig of 700 bp in length
42687 42786: gap of 100 bp
42787 43484: contig of 698 bp in length
43485 43584: gap of 100 bp
43585 44276: contig of 692 bp in length
44277 44376: gap of 100 bp
44377 45078: contig of 702 bp in length
45079 45178: gap of 100 bp
45179 45869: contig of 691 bp in length
45870 45969: gap of 100 bp
45970 46664: contig of 695 bp in length
46665 46764: gap of 100 bp
46765 47456: contig of 692 bp in length
47457 47556: gap of 100 bp
47557 48249: contig of 693 bp in length
48250 48349: gap of 100 bp
48350 49042: contig of 693 bp in length
49043 49142: gap of 100 bp
49143 49843: contig of 701 bp in length
49844 49943: gap of 100 bp
49944 50640: contig of 697 bp in length
50641 50740: gap of 100 bp
50741 51438: contig of 698 bp in length
51439 51538: gap of 100 bp
51539 52231: contig of 693 bp in length
52232 52331: gap of 100 bp
52332 53020: contig of 689 bp in length
53021 53120: gap of 100 bp
53121 53857: contig of 737 bp in length
53858 53957: gap of 100 bp
53958 54652: contig of 695 bp in length
54653 54752: gap of 100 bp

alignment_scores:

Quality: 284.00 Length: 80
Ratio: 4.733 Gaps: 2
Percent Similarity: 75.000 Percent Identity: 73.750

alignment_block:

US-09-135-238b-2 x AC027719/rev ..

Align seg 1/1 to reverse of: AC027719 from: 1 to: 58628

2 AspArgTrpLeuPrp..... 6

40676 GACGCGTGGAGCTGGGAAGTAGATGGGCTGGGATATAGTTAGATGGG 40627

7ProteuTyrPheLeuPro.....ValSerGlyAlaLeuArgI 19
|||||:|:| |
40626 GCTCCTCTCATTAATCTCCTCTCCCTGCAGTATCATGGGCGCCTGAGGA 40577
|
19 leLeuProGluValLysValGluGluGlyLeuGlySerValThrIle 35
|||||:|:| |
40576 TCTTCCCAAGAGTAAGGTAGAGGGAGCTGGGGGATCAGTTACCATC 40527
|
36 LysCysProLeuProGluMetHisValArgIleTyrLeuCysArgGluMe 52
|||||:|:| |
40526 AAGTGGCCACTCTCTGAAATGCATGTGAGGATATATCTGTGCCGGGAGAT 40477
|
52 talaclySerGlyThrCysGlyThrValValSerThrThr 65
|||||:|:| |
40476 GCTGGATCTGGAACATGTGGTACCGTGGTATCCACCACC 40437

seq_name: gb_ro:MMU06431

seq_documentation_block:

LOCUS MMU06431 3095 bp mRNA ROD 26-MAY-1995
DEFINITION Mus musculus B6/CBAF1J polymeric immunoglobulin receptor mRNA,
complete cds.
ACCESSION U06431
VERSION U06431.1 GI:458421
KEYWORDS
SOURCE mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3095)
AUTHORS Piskurich,J.F., Blanchard,M.H., Youngman,K.R., France,J.A. and
Kaetzel,C.S.

TITLE Molecular cloning of the mouse polymeric Ig receptor. Functional
regions of the molecule are conserved among five mammalian species
J. Immunol. 154 (4), 1735-1747 (1995)
MEDLINE 95138517
REFERENCE 2 (sites)
AUTHORS Piskurich,J.F., Hsieh,M., Youngman,K.R. and Kaetzel,C.S.

TITLE Molecular cloning of mouse polymeric Ig receptor
J. Immunol. 150, 38A-38A (1993)
REFERENCE 3 (bases 1 to 3095)
AUTHORS Piskurich,J.F.
TITLE Molecular Cloning and Regulation of the Polymeric Immunoglobulin
Receptor

Thesis (1994) Pathology, Case Western Reserve University
4 (bases 1 to 3095)
AUTHORS Piskurich,J.F.
TITLE Direct Submission
Submitted (08-FEB-1994) Janet F. Piskurich, Case Western Reserve
University, Pathology, Biomedical Research Building, Cleveland, OH
44060, USA

FEATURES

Location/Qualifiers
1..3095
/organism="Mus musculus"
/strain="B6/CBAF1J (C57 Black 6 X CBA)"
/db_xref="taxon:10090"
/clone="G20"
/clone_lib="catalog #935302, Stratagene Cloning Systems,
La Jolla, CA"
/sex="female"
/tissue_type="liver"
5'UTR <1..84
sig_peptide 85..138
CDS 85..2400
/codon_start=1
/function="binds and transports polymeric immunoglobulin"
/product="polymeric immunoglobulin receptor"
/protein_id="AAA67440.1"
/db_xref="GI:458422"
/translation="MRLLFTLLTVTFSGVSTKSPFGPQEVSSIEGDSVITCYTPD
TSVNRHRTKYWCROGASGCTTLISSNGYLSKEYSGRANLINFNNFTVNIQLQTO
DDTGSYKGLGTSNRGLSDVLEYSOYPLSDTHVYTKDIGNRVITIECFPRKNVP
SKKSLCKKTNQSCELYIDSTEVKNFSYIGRAKLFMKDLTIVTVYVNIHLTHNDAGLY

alignment_scores:

Quality: 284.00 Length: 80
Ratio: 4.733 Gaps: 2
Percent Similarity: 75.000 Percent Identity: 73.750

alignment_block:

US-09-135-238b-2 x AC027719/rev ..

Align seg 1/1 to reverse of: AC027719 from: 1 to: 58628

2 AspArgTrpLeuPrp..... 6

40676 GACGCGTGGAGCTGGGAAGTAGATGGGCTGGGATATAGTTAGATGGG 40627

```

|||||
561 TCCC 564

seq_name: gb_Om:AF091137

seq_documentation_block:
LOCUS AF091137 2848 bp mRNA MAM
DEFINITION Trichosurus vulpecula polymeric-immunoglobulin receptor precursor
(PiR) mRNA, complete cds.
01-JUL-1999

```

AF091137
AF091137.1 GI:5305512
Trichosurus vulpecula.

SOURCE	ORGANISM
Trichosurus vulpecula.	Trichosurus vulpecula
Trichosurus vulpecula	Trichosurus vulpecula

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.

REFERENCE 1 (bases 1 to 2848)

AUTHORS	Adamski, F.M. and Demmer, J.
TITLE	Cloning and characterization of pIGr and J chain of the marsupial, <i>Trichosurus vulpecula</i> (brush-tailed possum)

JOURNAL
Unpublished
Strenuous variegata (brush: called possum)

REFERENCE
2 (bases 1 to 2848)
AUTHORS
Adamski, F.M. and Demmer, J.

TITLE Direct Submission
JOURNAL Submitted (09-SEP-1998) Dairy Science, AgResearch Ruakura, East

FEATURES

Street, Hamilton, New Zealand

Location/Qualifiers

```
source
1. .2848
/organism="Trichosurus vulpecula"
```

```

2020-01-20 10:00:00 /db_xref="taxon:9337"
2020-01-20 10:00:00 /tissue_type="mammary gland"

```

```

gene
T. .2848
/gene="PIGR"
110 250
chr

```

```

CDS
145..12550
/gene="PICR"
/codon_start=1

```

```

/condition_start=1
/product="polymeric-immunoglobulin receptor precursor"
/protein_id="AAD41688.1"

```

```

/ProteinData/BankIt1000001
/db_xref="GI:5305513"
/translation="NAFFLACLALLPVVSMKSPLEFGPKVOTGVGGSSVSIOTGYPSS

```

SYNHRGRYFCLQNRQSCETIVSSNGFVSERFSGRAKLTNFPNGNSFLIQISLEK
DIGLYKCGLGTSNRLGSFDITLVEVGEGPNLPNNTEIVITEVGKTVSINCPFOBQNTQ

RFELCKKDGESCALVIDSQEQVGPDYTGRRRLSISGTSRRFVVVTISQIKRQDVGMY
CGVGEDSDTGIQKNVDLKLLEPEPELLYAEELGGSVTLNLCALGSTWASVPKFLCQMR

ETCDLVINSKGFTNNATHGRILFSHTTTETSGFSIMITQVRKEDEGVYHCGAQEDGQFP
EEGPIRALQLFVSEETTPVPSPLVWKGSGGSVTITCHYDPKKNNTLKYWCWKEGSSS

CTKLVDLSLGMVDESIEYGRVALWDEPENGIFTVILNQLTPQDAGYYWCLSNGEHNRKS
VKIEINDGQPLLIAPKTVTAQLGQSLTISCHFPCKFYSEYKWKSNQGCETLPTQ

EGSSQAFPCNQNSRNSLTLSNSVTRDHEGWYWCQVKNQNGYGETTAVSVSVSEEEVS
 NAIQPTNAVLNEDAVEPKVRGKEIEVPTDLGSTEEHSGGSVLVSTLPPLALVTGVGG
 NAIQPTNAVLNEDAVEPKVRGKEIEVPTDLGSTEEHSGGSVLVSTLPPLALVTGVGG

760 600 700 800
IITQTNPEIIQQPKVQRSSKEEDMAYTAFLLQSNMMANEIPLDOPSNA"
VALGILRRKRWFSDRKVSQSYRTDLSELENPNRQFGANENMDASVQETTLGGED

BASE COUNT	769 3	690 C	722 g	667 f
ORIGIN				

alignment scores.

argument_scores:		
Quality:	192.00	Length: 204
Ratio:	1.684	Gans: 8

Percent Similarity: 55.882 Percent Identity: 30.392

```
alignment_block:
US-09-135-238B-2 x AF091137
:
```

Align seg 1/1 to: AF091137 from: 1 to: 2848

17 LeuArgIleLeuProGluVal..... 23

173 CTGCTCTCTCTCCCTGGTTCTATGAAGAGTCCCATATTTGGCCCCAA 222

24 .LysValGluGlyGluLeuGlyGlySerValThrIleLysCysProLeup 40

223 GCAAGTGACAGGAGTGGAAGGAGGATGTGTCTCCATCCAGTGCTTCTACC 272

100

/gene="pIgR"
/label=ex3
/usedin=Y16524:pig_cds
/usedin=Y16524:pig_mrna
BASE COUNT 130 a 136 c 127 g 112 t
ORIGIN

alignment_scores:
Quality: 188.00 Length: 97
Ratio: 2.648 Gaps: 3
Percent Similarity: 73.196 Percent Identity: 40.206
alignment_block:
US-09-135-238B-2 x MPMIGR3 ..
Align seg 1/1 to: MPMIGR3 from: 1 to: 505

30 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 42
||| |||||:|||| ||| |||||
130 GCGGACTCTGTTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 179
42 ThisValArgIleTyLeuCysArgGluMetAlaGlySerGlyThrCysG 59
|||:||||| ||| |||||: ||| ||||| |||
180 GCACACCCGGAATACTGGTCCGACAAAGGAGCC....AGCGGCATGTGCA 226
59 lyThrValValSerThrThrAsnPhelLysAlaGluTyLysGlyArg 75
|||:|||||:|||||:|||||: |||||: |||||
227 CAACGCTCATCTCTCAATGGCTACCTCTCCAGGAGTATTACGGCAGA 276
76 ValThrLeuLysGlnTyProArgLysAsnLeuPheLeuValGluValTh 92
:|||||: |||||: |||||: |||||: |||||: |||||
277 GCCAACCTCATCACTCCAGAGAACACACATTTGTGATTAAACATTGA 326
92 rGlnLeuThrGluSerAspSerGlyValTyAlaCysGlyAlaGlyMeta 109
:|||||: |||||: |||||: |||||: |||||: |||||
327 GCAGCTACCCAGGAGCTGCTTCGATGTCAGCCTGGAGGTC 414
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
:|||||: |||||: |||||: |||||: |||||: |||||
374 CCAGTAACCGAGGCTGCTCCTTCGATGTCAGCCTGGAGGTC 414

seq_name: gb_pat:E15253

seq_documentation_block:
LOCUS E15253 29392 bp DNA PAT 28-JUL-1999
DEFINITION Mouse pig receptor gene.
ACCESSION E15253
VERSION E15253.1 GI:5709936
KEYWORDS JP 1998057066-A/1.
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Sato,T. and Kushiroy,A..
TITLE MURINE PIG RECEPTOR PROTEIN GENE
JOURNAL Patent: JP 1998057066-A 03-MAR-1998;
YAKULT HONSHA CO LTD
COMMENT OS Mus sp. (mouse)
PN JP 198057066-A/1
PD 03-MAR-1998
PF 19-AUG-1998 JP 1996217154
PI SATO TAKASHI, KUSHIRO AKIRA
PC C12N15/09,C07H21/04,C07K14/705;
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FH 1. .29392
FT source
FT /organism='Mus sp.'
FT /clone_lib='genomic DNA from female 129SVJ FT
FT mouse liver in

FEATURES
source Location/Qualifiers
1..29392
/organism="Mus sp."
/db_xref="taxon:10095"
BASE COUNT 8318 a 6747 c 6514 g 7813 t
ORIGIN

alignment_scores:
Quality: 188.00 Length: 97
Ratio: 2.648 Gaps: 3
Percent Similarity: 73.196 Percent Identity: 40.206

alignment_block:
US-09-135-238B-2 x E15253 ..
Align seg 1/1 to: E15253 from: 1 to: 29392
30 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 42
||| |||||:|||| ||| |||||
18245 GCGGACTCTGTTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 18294
42 ThisValArgIleTyLeuCysArgGluMetAlaGlySerGlyThrCysG 59
|||:||||| ||| |||||: ||| ||||| |||
18295 GCACACCCGGAATACTGGTCCGACAAAGGAGCC....AGCGGCATGTGCA 18341
59 lyThrValValSerThrThrAsnPhelLysAlaGluTyLysGlyArg 75
|||:|||||:|||||:|||||: |||||: |||||
18342 CAACGCTCATCTCTCAATGGCTACCTCTCCAGGAGTATTACGGCAGA 18391
76 ValThrLeuLysGlnTyProArgLysAsnLeuPheLeuValGluValTh 92
:|||||: |||||: |||||: |||||: |||||: |||||
18392 GCCAACCTCATCACTCCAGAGAACACACATTTGTGATTAAACATTGA 18441
92 rGlnLeuThrGluSerAspSerGlyValTyAlaCysGlyAlaGlyMeta 109
:|||||: |||||: |||||: |||||: |||||: |||||
18442 GCAGCTACCCAGGAGCAGCTGGGAGCTACAAGTGTGCGCTGGGT...A 18488
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
:|||||: |||||: |||||: |||||: |||||: |||||
18489 CCAGTAACCGAGGCTGCTCCTTCGATGTCAGCCTGGAGGTC 18529

seq_name: gb_to:AB001489

seq_documentation_block:
LOCUS AB001489 29392 bp DNA ROD 14-APR-2000
DEFINITION Mus musculus DNA for polyimmunoglobulin receptor, complete cds.
ACCESSION AB001489
VERSION AB001489.1 GI:2804245
KEYWORDS polyimmunoglobulin receptor.
SOURCE Mus musculus (Strain:129SVJ) female liver DNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Kushiroy,A. and Sato,T.
TITLE PolymERIC immunoglobulin receptor gene of mouse: sequence,
JOURNAL structure and chromosomal location
MEDLINE Gene 204 (1-2), 277-282 (1997)
REFERENCE 98094283
AUTHORS Sato,T.
TITLE Direct Submission
JOURNAL Submitted (26-FEB-1997) to the DDBJ/EMBL/GenBank databases. Takashi
Sato, Yakult Central Institute for Microbiological Research,
Culture Collection Center; 1796 Yaho, Kunitachi-shi, Tokyo 186,
Japan (Tel:425-77-8960, Fax:425-77-3020)
FEATURES
source Location/Qualifiers
1..29392
/organism="Mus musculus"
/strain="129SVJ"
/db_xref="taxon:10090"


```

KEYWORDS  JP 1998072495-A/1.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens.
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS    Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      1 (bases 1 to 1911)
JOURNAL    Miyabayashi,T. and Sakano,S. .
           IMMUNITY-RELATED FACTOR
Patent: JP 1998072495-A 17-MAR-1998;
ASNTI CHEM IND CO LTD
OS         Homo sapiens (human)
PN         JP 1998072495-A/1
PD         17-MAR-1998
PR         11-JUN-1997 JP 1997153218
PR         13-JUN-1996 JP 96P 152362
PI         MIYABAYASHI TOMOYUKI, SAKANO SEIJI
PC         C07K14/47,C07H21/04,C07K16/18,C12N5/10,C12N15/09, PC
C12P21/02//A61K38/00,
PC         C12P21/08,(C12N5/10,C12R1:91),(C12P21/02,C12R1:91),(C12P21/08,
PC         C12R1:91);
CC         strandedness: Double;
CC         topology: Linear;
CC         hypothetical: No;
CC         anti-sense: No;
FH         Key      Location/Qualifiers
FH         1. .1911
FT         source    /organism='Homo sapiens'
FT         /tissue_type='Lymph node'
FT         5'UTR     1. .97
FT         sig_peptide 98..145
FT         /product='signal peptide of immunity related
FT         protein'
FT         mat_peptide 146..1693
FT         /product='immunity related protein' FT   CDS
FT         98..1696 /product='immunity related protein' FT   3'UTR
FT         1697..1911.
FEATURES    Location/Qualifiers
            source    1. .1911
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT 490 a 541 c 525 g 355 t
ORIGIN

alignment_scores:
  Quality: 197.50      Length: 575
  Ratio: 0.906         Gaps: 20
  Percent Similarity: 36.000 Percent Identity: 18.957

alignment_block:
US-09-135-238B-2 x E15470 ..
Align seg 1/1 to: E15470 from: 1 to: 1911
3 ArgTrpLeuTrpProLeuTyPheLeuProValSerGlyAlaLeuArgI1 19
|||||
170 AGATGGCTGTGGGAG...GGCTCTCTCCCTCCAGGACCCCATCTCCGGGC 216
19 eLeuProGluValLys.....
:::
217 CATGGGAACACTCAGGCCTTCTCGCCCTCTGCTGGGGGAGGAGAGCT 266
25 .....
267 COTTTGCAGCTCCAAATTCAATTGAAGGCTCAAGGCTGGTGTGTCAGGGAG 316
29 LeuGlyGlySerValThrIleLysCysProLeuPro.....G1 41
|||||
317 CTGGAGGAGCTGTACCATCATCGTGCATATATGCCCCCTCATCTGTCAA 366
41 uMethHisValArgIleTyLeuCysArgGluMetAlaGlySerGlyThrC 58

```

```

::: ||| ||| ||| ||| |||
367 CAGGCACACGAGGAAGTACTGGTGGCGTCTGGGGCCCCCAAGATGATCT 416
58 ySGlyThrValValSerThrThrAsnPheIleLysAlaGluTyLysGly 74
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
417 GCCAGACCATGTGTCCACCAACCAAGTATATCTACCATCGCTATCGTGAC 466
75 ArgValThrLeuLysGlnTyProArgLysAsnLeuPheLeuValGluVa 91
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
467 CGTGTGGCCCTCACAGACTTTCACAGAGAGAGGCTTGTGTGGTGAGGCT 516
91 lThrGlnLeuThrGluSerAspSerGlyValTyAlaCysGlyAlaGlyM 108
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
517 GTCCCAACTGTCCCGGATGACATCGGATGCTACCTCTCGGCGCATTTGAA 566
108 etAsnThrAspArg.....
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
567 GTGAAAACAACATGCTGTTCTTAAGCATGNAATCTGACCATCTCTCGAG 616
112 .....
617 CCGCCAGCACCTCCCCACAGCCACTCCAGCTGCTGGGAGCTCACCAT 666
113 .....
667 GAGATCCTATGGAACAGCGTCTCCAGTGGCCAAACAGATGGACCCCAAGAA 716
114 yThrGln.....
|||||
717 CCACCCAGACCTTAGACAGGGGACAGCATGGGACACAGTGTCTTCCACT 766
117 .....LysValThrLeuAsnValHisSerGluTyTrpGluPro... 128
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
767 CCAGGAACCAAGCAAGACTACAGCTCAGCTGAGGGAAGACAGCAACCCAGG 816
129 .....SerTrpGluGlu..... 132
817 AGCAACACGAGCCAGCAGCTCCAGGGACAGCAGCTGGGCAGAGGGTCTCG 866
133 .....GlnProMetProGluThrProLysTrpPheHisLeuPro 145
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
867 TCAAAGCAGCTGTCTCGGATTCCAGAGAGTCCA..... 898
146 TyrLeuPheGlnMetProAlaTyAlaSerSerLysPheValThrAr 162
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
899 .....COTTCAAAGACGAGAAGCATGTCCAA 924
162 gValThrThrProAlaGlnArgGlyLysValProValHisHisSerS 179
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
925 TACAACAGAAAGGTGTGGGAGGC..... 949
179 erProThrThrGlnIleThrHisArgProArgValSerArg..... 192
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
950 ..ACCAGAAGCTCGGTGACAAACAGGGCTAGAGCCAGCAAGGACAGGAGG 997
193 ...AlaSerSerValAlaGlyAspLysProArg..... 202
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
998 GAGATGACAACTACCAAGGCTGATAGGCCAAGGAGGACATAGAGGGGT 1047
202 ..... 202
1048 CAGGATAGCTCTTGATGACGCCAAAAGGTCTTAGAACCATTTGGGCCAC 1097
203 .....ThrPheLeuProSerThrThr 209
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1098 CAGCTCTGGTCTCAGAAACTTTGGGCTGGGAAATCTCCCAACAAGCAAG 1147
210 AlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProGlnThrProSe 226
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1148 CCAGTTTCTTAAGCAACAATCTCAGGGTTCCATTGGGAAACAACATCCAGC 1197
226 rTyraHisHisThrArg..... 232
::: |||

```

```

1198 TGCAGCATGTGGACCTTGGAACTCCAGCTGCAGATGTGTGGATCTTGG 1247
233 .....LeuHisArgGlnArgAlaLeuAspTyr 241
1248 GAACCTCCAGCTGCAGATGTGTGACAGCATGGAGCAGCATCTGGGAA 1297
242 GlySerGlnSerGly..... 246
1298 GGAAGCGCTGCAGGGGACCTAGATGCTGCCACTGGACAGAGGTCCCCA 1347
246 ..... 246
1348 AGCAACACTGAGCCAGACCCCGCAGTAGAGCCCTGGGGACCCCTGGCA 1397
247 .....ArgGluGlyGlnGlyPheHis 253
1398 AGGAGTCCCTCGTGAAGCGTACTTTTCCAGAAGATGAAGACGACTCTCGG 1447
254 IleLeuIlePro.....ThrIleLeuGlyLeuPheLeuAlaLeuLe 268
1448 ACCTGCTGCTGCTCTTACCATGCTGGCCCTGTTTATGCTTATGGCTCT 1497
268 uGlyLeuValValLysArgAlaValGluArgAlaLeuSerArgA 285
1498 GTTCTATTG.....CAAAGGAAGCTCTGGAGAAGGA 1529
285 rGalaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln... 300
1530 GGACCTCTCAGAGGCGAAGAGGTCACTTAATTCAGATGACACATTTT 1579
301 .....ArgProArgGlySerProArgProArgSerGlnAsnIle 314
1580 CTGGAAGTAGAACCCCAAGCAGACACAGCTGCCCATGTGGAAAGAAAGAT 1629
314 eTyr.....SerAlaCysProArgArgAlaArgGlyAlaAspAlaAag 329
1630 GCTCCAGGATGACTCTCTTCT.....GCTGGGCGCAGCC 1664
329 lyThrGlyGluAlaProValProGlyPro.....Gly 339
1665 TGACTGCCCCAGAGAAATCCAGGACCTGAGGACAGACAGATGAACCT 1714
340 AlaProLeuPro.....ProAl 345
1715 GCTCAGTTACCTGGAGAGGACCAAGATCAAGGCTTCAGGACCCCA 1764
345 aProLeuGlnValSerGluSerPro 353
1765 GCCTCTTTCCATCATCTCTCTCCA 1789

seq_name: gb_ro:RNIGR

seq_documentation_block:
LOCUS RNIGR 3269 bp mRNA ROD 22-MAR-1995
DEFINITION Rat mRNA for polymeric immunoglobulin receptor.
ACCESSION X15741
VERSION X15741.1 GI:56464
KEYWORDS immunoglobulin receptor; polymeric immunoglobulin receptor;
transmembrane protein.
SOURCE Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 3269)
AUTHORS Banting,G.
TITLE Direct Submission
JOURNAL Submitted (05-JUL-1989) Banting G., EMBL, Meyerhofstr 1, 6900
Heidelberg
REFERENCE
2 (bases 1 to 3269)
AUTHORS Banting,G., Brake,B., Braghetta,P., Luzzio,J.P. and Stanley,K.K.
JOURNAL Unpublished
FEATURES
Location/Qualifiers
1..3269 source

```

```

/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_type="liver"
/cloned_lib="CDNA_TRL1"
/cloned="GORE1,8,43,4339"
74..127
/product="rat polymeric immunoglobulin receptor signal
peptide (AA -18 to -1)"
74..2383
/notes="rat polymeric immunoglobulin receptor (AA -18 to
751)"
/codon_start=1
/protein_id="CAA33758.1"
/db_xref="GI:56465"
/db_xref="SWISS-PROT:P15083"
/translation="MRLSLFALLVTVFSGVTSQSPFQPDVSSIEGNSVSTCYYPD
TSVNRHTRKYWCROGANGCATLISNGYLSKEYSGRASLINFENSTFVNIHLTO
EDTSYKGLGTNRGLFDVSVLEVSQVPEFNDTHVTKDIGRTVTIECRKEGNAH
SKSLCKKSGEACEVVIDSTYVDPYSKDRALLFMKGTSRDIFVNIHLIPSDAGLY
VCQAGESPADKNADLOVLEPEPELLYKDLSSVTPECDLGREVANDAKYLCRKKE
TCDVILNLTGRDPAFEGRIILLTPDDNGRFSVLITGRKEDAGHYQCAHSSGLPQE
GWPQAWQLFVNEESTIPNSRVVKGVTGSVAIVCPYNPRESSSLKYWCHEADENG
RCPLYVQALYQOEGYEGRLALDPQSGAYTVILNQLTQDSGFYCLTDGDSKWT
TIELOVAEATKKPDLVTPQNATVIGETFTISCHYPCFKYQKWCXNSDGHIL
PSHDEGARQSSVCDSSQIVMTLPNPKKDEGWYKCVGKQGVYGETTAIYVAEE
RTGSGPHINTDANARAKDAPPEEAMESSVREDENKANLDPRLFADERIONAGDQAO
ENRASNAGSAGSGSSKVLFTLVLPLGLVLAGAVAVVAVRVHRKKNVDRMSISSY
RTDISMGDFRNSRDLGGNDNMGATPDQETVLEGKDELETTTCTTEPEESKKARRS
KEADAMAYSAFLFQSSTTAQVHDGQEA"
128..2380
mat_peptide /product="rat polymeric immunoglobulin receptor mature
peptide (AA 1 to 751)"
BASE COUNT 890 a 831 c 850 g 698 t
ORIGIN

alignment_scores:
Quality: 185.50 Length: 164
Ratio: 1.855 Gaps: 6
Percent Similarity: 60.976 Percent Identity: 30.488

alignment_block:
US-09-135-238b-2 x RNIGR ..
Align seg 1/1 to: RNIGR from: 1 to: 3269

30 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 42
|||||:|||||:|||||:|||||:
170 GGTAACCTCGGTCTCCATCAGCTGCTACTACCAGACACCTCTGTCAACCG 219

42 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
|||||:|||||:|||||:|||||:
220 GCACACCCGGANAATATCTGTCGCCACAGGAGCC...AACGGCTACTGCG 266

59 lyThrValValSerThrAsnPheIleLysAlaGluTyrLysGlyArg 75
:::|||||:|||||:|||||:|||||:
267 CAACCTCATCTCTTCAATGGCTACCTCTCGAAGGAGTATTCAGGCAGA 316

76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92
:::|||||:|||||:|||||:|||||:
317 GCACCTCATCACTTCCAGANATGACATTTGTGATTACATGCG 366

92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 109
:::|||||:|||||:|||||:|||||:
367 ACATCTCACCCAGGAGGACACTGGGAGCTACAAAGTGTGCTGGGT...A 413

109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
:::|||||:|||||:|||||:|||||:
414 CCACTAACCCAGGCGCTGTTTTTCGATGTCAGCCTGGAGGTC..... 454

126 TyrGluProSerTrpGluGlnProMetProGluThrProLysTrpPh 142
:::|||||:|||||:|||||:
455 .....AGCCAGGTTCTCTGAGTTCCTCCAAATGACAC 483

```


LOCUS S43435S03 345 bp DNA PRI 08-MAY-1993
DEFINITION transmembrane secretory component [human, leukocytes, Genomic, 345
nt, segment 3 of 11].
ACCESSION S43441
VERSION S43441.1 GI:255089
KEYWORDS 3 of 11
SEGMENT human leukocytes.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota: Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 345)
AUTHORS Krajci,P., Kvale,D., Tasken,K. and Brandtzaeg,P.
TITLE Molecular cloning and exon-intron mapping of the gene encoding
human transmembrane secretory component (the poly-Ig receptor)
JOURNAL Eur. J. Immunol. 22 (9), 2309-2315 (1992)
MEDLINE 92387236
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 11232] from the original journal article.
This sequence comes from Fig. 2.

FEATURES
source Location/Qualifiers
1..345 /organism="Homo sapiens"
78 a 102 c 97 g 68 t /db_xref="taxon:9606"

BASE COUNT 78 a 102 c 97 g 68 t
ORIGIN

alignment_scores:
Quality: 182.00 Length: 115
Ratio: 2.247 Gaps: 5
Percent Similarity: 70.435 Percent Identity: 37.391

alignment_block:
US-09-135-238B-2 x S43435S03 ..
Align seg 1/1 to: S43435S03 from: 1 to: 345

12 ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGly 28
||||| ||| ||||| :|||:|||||
18 CCATATTGGT.....CCCGAG...GAGGTGAATAGTGT 49
28 uLeuGlyGlySerValThrIleLysCysProLeuPro.....G 41
|||||:|||||:||||| ||| :|||
50 GGAAGGTAACCTCAGTGTCCATCAGCTGCTACTACCCACCCACCTCTGTCA 99
41 luMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThr 57
:: |||:|||| ||| |||||: ||| :|||
100 ACCGGCACACCCGGAAGTACTGGTCCGCGGAGGAGCTAGAGGTGC... 146
58 CysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysG1 74
||| |||:|||||: ||| :|||:|||||: ||| |||
147 TGCATAACCCCTCATCTCTCGGAGGGCTACGTCCTCCAGCAATATGCAGG 196
74 YArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluV 91
:|||||:|||||: |||:|||||: |||:|||||: |||
197 CAGGGCTAACCTCACCACCTCCCGGAGAACGGCACATTTGTGGTGAACA 246
91 alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107
:|||||:|||||: |||:|||||: ||| ||||| |||
247 TTGCCCAGCTGAGCCAGGATGACTCCGGCGGCTACAAGTGTGGCTGGGC 296
108 MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
:|||||: ||| ||| :|||:|||||: |||
297 ATCATATAGC...CGAGGCTGTCCTTTGATGTCAGCCTGGAGGTC 338

seq_name: gb_pat:A52094

seq_documentation_block:
LOCUS A52094 1936 bp DNA PAT 11-MAR-1997
DEFINITION Sequence 4 from Patent WO9618734.
ACCESSION A52094
VERSION A52094.1 GI:2304706

KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1936)
AUTHORS Van,O.J., Asselbergs,F.A., Leist,C. and Gruetter,M.
TITLE PRODUCTION OF RECOMBINANT SECRETORY COMPONENT
JOURNAL Patent: WO 9618734-A 4 20-JUN-1996;
CIBA GEIGY AG (CH)
COMMENT Other publication AU 4302796 960703.
FEATURES Location/Qualifiers
source 1..1936 /organism="unidentified"
mat_peptide 16..1920 /db_xref="taxon:32644"
BASE COUNT 451 a 521 c 583 g 381 t /product="TRANSCRIPT FROM SC DNA FRAGMENT 2"
ORIGIN

alignment_scores:
Quality: 182.00 Length: 115
Ratio: 2.247 Gaps: 5
Percent Similarity: 70.435 Percent Identity: 37.391

alignment_block:
US-09-135-238B-2 x A52094 ..
Align seg 1/1 to: A52094 from: 1 to: 1936

12 ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGly 28
||||| ||| ||||| :|||:|||||
76 CCATATTGGT.....CCCGAG...GAGGTGAATAGTGT 107
28 uLeuGlyGlySerValThrIleLysCysProLeuPro.....G 41
|||||:|||||:||||| ||| :|||
108 GGAAGGTAACCTCAGTGTCCATCAGCTGCTACTACCCACCCACCTCTGTCA 157
41 luMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThr 57
:: |||:|||| ||| |||||: ||| :|||
158 ACCGGCACACCCGGAAGTACTGGTCCGCGGAGGAGCTAGAGGTGC... 204
58 CysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysG1 74
||| |||:|||||: ||| :|||:|||||: ||| |||
205 TGCATAACCCCTCATCTCTCGGAGGGCTACGTCCTCCAGCAATATGCAGG 254
74 YArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluV 91
:|||||:|||||: |||:|||||: |||:|||||: |||
255 CAGGGCTAACCTCACCACCTCCCGGAGAACGGCACATTCGTGGTGAACA 304
91 alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107
:|||||:|||||: |||:|||||: ||| ||||| |||
305 TTGCCCAGCTGAGCCAGGATGACTCCGGCGGCTACAAGTGTGGCCTGGGC 354
108 MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
:|||||: ||| ||| :|||:|||||: |||
355 ATCAATAGC...CGAGGCTGTCCTTTGATGTCAGCCTGGAGGTC 396

```

121 TTTTCAGGGGCTCCACAAAGGCCCATATTTGGTCCCGAGGAGTGAG 170
122 TTTTCAGGGGCTCCACAAAGGCCCATATTTGGTCCCGAGGAGTGAG 170
26 uGlyGluLeuGlySerValThrIleLysCysProLeuProGlu... 41
171 TAGTATAGAGGCGACTCTGTTCCATCAGCTGCTACTACCGAGACACT 220
42 .....MetHisValArgIleTyrLeuCysArgGluMetAlaGlySer 55
221 CTGTCAACCGGCACACCCGGAATACTGGTGGCCGACAGGAGCC...AGC 267
56 GlyThrCysGlyThrValValSerThrThrAsnPhelLysAlaGluTyr 72
268 GCATGTCGCAACGCTCATCTCTCAATGGCTACTCTCCAAAGAGTA 317
72 rlyGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuV 89
318 TTCAGGAGAGCAACCTCATCACTTCCGAGAGAACACACATTTGTGA 367
89 aGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGly 105
368 TTAACATTGAGCAGCTACCCAGGAGCAGACTGGGAGCTACAAGTGTGSC 417
106 AlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVa 122
418 CTGGGT...ACCAGTAACGAGGCGCTCTCCGATGTCAGCTGGAGGT 464
122 lHisSerGluTyrGluProSerTyrGluGlnProMetProGluThrP 139
465 C.....AGCCAGTTCCTGAGTTCG 484
139 rOlyTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSer 155
485 CGAGTGACACCCAGCTC.....TACACAAAG 510
156 SerSerLysPheValThrArgValThrProAlaGlnArgGlyLysVa 172
511 GACATAGCAGAAATGTACCATTTGATGCGCTTTCAAAGAGGAGAAATGT 560
172 lPro 173
561 TCCT 564

seq_name: /cgn2_6/ptodata/1/ina/5_comb.seq:us-08-434-000A-9

seq documentation block:
: Sequence 9, Application US/08434000A
: Patent No. 6046037
: GENERAL INFORMATION:
: APPLICANT: ANDREW C. HIATT, JULIAN
: APPLICANT: K.-C. MA, THOMAS LEHNER
: TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: 633 West Fifth Street
: STREET: Suite 4700
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90071
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: MEDIUM TYPE: storage
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: IBM P.C. DOS 5.0
: SOFTWARE: Word Perfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/434,000A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:

```

```

: PRIOR APPLICATION DATA: including application
: PRIOR APPLICATION DATA: described below: 1
: APPLICATION NUMBER: 08/367,395
: FILING DATE: 12/30/94
: ATTORNEY/AGENT INFORMATION:
: NAME: Guise, Jeffrey W.
: REGISTRATION NUMBER: 34,613
: REFERENCE/DOCKET NUMBER: 212/127
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 552-8400
: TELEFAX: (619) 552-0159
: TELEX: 67-3510
: TELETYPE: SEQUENCE LISTING
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3269 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: DESCRIPTION: Rat Polyimmunoglobulin Receptor
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 74.....2383
: US-08-434-000A-9

alignment_scores:
: Quality: 185.50 Length: 164
: Ratio: 1.855 Gaps: 6
: Percent Similarity: 60.976 Percent Identity: 30.488

alignment_block:
US-09-135-238B-2 x US-08-434-000A-9 ..
Align seq 1/1 to: US-08-434-000A-9 from: 1 to: 3269
30 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 42
|||||
170 GGTAACTCGGTCTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACGG 219
42 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
|||||
220 GCACACCCGGAATCTACTGGTCCGACAGAGGCC...AAGGGCTACTGGG 266
59 lYThrValValSerThrThrAsnPhelLysAlaGluTyrLysGlyArg 75
|||||
267 CAACCTCTCTCTCAATGGCTACTCTCGAAGAGTATTCAGCAGA 316
76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92
|||||
317 GCACGCTCATCACTACCTCCGAGAGAAATAGCAGATTGTGATTAACATTGC 366
92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 109
|||||
367 ACATCTCACCCAGGAGGACACTGGGAGCTACAAGTGTGGTGGGT...A 413
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
|||||
414 CCACCTAACCCGAGGCGCTGTTTTCGATGTCAGCTGGAGTGC..... 454
126 TyrGluProSerTyrGluGlnProMetProGluThrProLysTrpH 142
455 .....AGCCAGTTCCTGAGTTCCTCAATACACAC 483
142 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 159
|||||
484 CCAATGC.....TACACAAAGG 500
159 heValThrArgValThrThrProAlaGlnArgGlyLysValProProVal 175
|||
501 ACATAGGCAAGAACTGTGACCATTCGAATGCCGTTTCAAAGAGGAGGAATGCT 550
176 HisHisSer.SerProThrThrGlnIleThrHisArgPro 188

```

```

111  ....:||||| .....: :|||
551 CATAGCAAGAAATCCCTGTGTAAAGAGAGAGAGGCCT 590

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-3

seq_documentation_block:
; Sequence 3, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2919 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 235.....2472
US-08-434-000A-3

alignment_scores:
Quality: 182.00 Length: 115
Ratio: 2.247 Gaps: 5
Percent Similarity: 70.435 Percent Identity: 37.391

alignment_block:
US-09-135-238b-2 x US-08-434-000A-3 ..
Align seg 1/1 to: US-08-434-000A-3 from: 1 to: 2919

12 ProValSerGlyAlaLeuArgIleLeuProGluValGlyGly 28
||||| :|||
241 CCCATATTGGT.....CCCGAG...GAGGTGAATAGTGT 272
```

```

28 uLeuGlyGlySerValThrIleLysCysProLeuPro.....G 41
|||||:|||||:||||| ||| |||
273 GGAAGTAACTAGTGTCCATCAGTGTACTACCCACCCACCTGTGCA 322
41 luMethisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThr 57
:::|||||:|||||:|||||:|||||
323 ACCGGCACACCCCGAAGTACTGGTCCCGCAGGAGCTAGAGTGGC... 369
58 CysGlyThrValSerThrThrAspPheIleLysAlaGluTyrLysGly 74
|||||:|||||:|||||:|||||:|||||:|||||
370 TGCATACCTCATCTCTCCGAGGCTAGCTCTCCAGCAATATGCAGG 419
74 YATGValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 91
|||||:|||||:|||||:|||||:|||||:|||||
420 CAGGGCTAACTCACCACACTTCCCGAGAGACGGCACATTTGTGTGAACA 469
91 alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107
:::::|||||:|||||:|||||:|||||:|||||
470 TTGCCAGCTGAGCCAGGATGACTCCGGCGGCTACAAGTGTGGCCTGGGC 519
108 MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
:::|||||:|||||:|||||:|||||:|||||:|||||
520 ATCAATAGC...CGAGGCTGTCTTGTGATGTGAGCTGAGGTC 561

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-5

seq_documentation_block:
; Sequence 5, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3630 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```

; TOPOLOGY: DESCRIPTION: Bovine Polymunoglobulin Receptor
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 152.....2425
; US-08-434-000A-5

alignment_scores:
  Quality: 182.00      Length: 123
  Ratio: 2.193        Gaps: 4
  Percent Similarity: 67.480      Percent Identity: 34.146

alignment_block:
  US-09-135-238B-2 x US-08-434-000A-5 ..
  Align seg 1/1 to: US-08-434-000A-5 from: 1 to: 3630
    17 LeuArgIleLeuProGluValIysValIleGluLeu..... 29
    179 CTGGCCATCTTCCAGTGGTCTCCATGAAGAGTCCCATCTTCGGTCCCGA 228
    30 .....GlyGlySerValThrIleLysCysProLeu 40
    229 GGAGGTGAGCAGCGTGGAGCGCGCTCAGTGTCCATCAAGTGTACTACC 278
    40 ro.....GluMetHisValArgIleTyrLeuCysArgGluMet 52
    279 CGCCACCTCCGTCAACCGCGCACGCGCAAGTACTGTGCGCGCAGGGA 328
    53 AlaGlySerGlyThrCysGlyThrValValSerThrAsnPheIleLeu 69
    329 GCC...CAGGCGCGCTGCACGACCTCATCTCCTCGAGGCGTACGTCTC 375
    69 sAlaGlyTyrIysGlyArgValThrLeuLysGlnTyrProArgIysAsnL 86
    376 CGAGCACTACTGGCGCAGAGCAACCTCACCACCTCCCGAGAGCGGCA 425
    86 eupHeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyr 102
    426 CGTTGTGTGGATCATCAGCCATCTCACCATAAAGACTCAGGCGCTAC 475
    103 AlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValTh 119
    476 AGTGTGCGCTGGCATATGACGCG...CGTGGCTTAATCTCGATGTGAG 522
    119 rHeuAsnValHisSerGlu 125
    523 CTGGAGGTACGCAAGAT 541

seq_name: /cgn2_5/ptodata/1/ina/5D_COMB.seq:US-08-642-406A-21

seq_documentation_block:
; Sequence 21, Application US/08642406A
; Patent No. 5959177
; GENERAL INFORMATION:
; APPLICANT: Heintz, Mich B.
; APPLICANT: Hiatt, Andrew C.
; APPLICANT: Ma, Julian K.C.
; TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED
; TITLE OF INVENTION: SECRETORY ANTIBODIES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
; STREET: 10666 No. 5959177th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,406A
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/591,823
; FILING DATE: 02-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/427,765
; FILING DATE: 27-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Logan, April C.
; REGISTRATION NUMBER: 33,950
; REFERENCE/DOCKET NUMBER: 184.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1875 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1875
; US-08-642-406A-21

alignment_scores:
  Quality: 166.50      Length: 103
  Ratio: 2.602        Gaps: 4
  Percent Similarity: 62.136      Percent Identity: 42.718

alignment_block:
  US-09-135-238B-2 x US-08-642-406A-21 ..
  Align seg 1/1 to: US-08-642-406A-21 from: 1 to: 1875
    30 GlyCysSerValThrIleLysCysProLeuPro.....GluMe 42
    115 GGGGACTCGGTGCTCCATCATCTACTACCAACCACTCCGTCACCCG 164
    42 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
    165 GCACAGCCGGAAGTCTGTGTCCGGGAA...GAGGAGAGCGCGCTGCG 211
    59 lyThrValValSerThrThrAsnPheIleLysAlaGlyTyrLysGlyArg 75
    212 TGACGCTTGCCTCG...ACGGCTACAGTCCCAAGCAATCTCCGGGAGA 258
    76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92
    259 GGCACACTCAGGACTTCCTCCATAGAGGAGGAGTTGGTGGTACTGTGA 308
    92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 109
    309 CCACACTCACCCAGACGACTCAGGAGGAGCTACAAGTGTGCGTGGGAGTCA 358
    109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
    359 AC...GGCGGTGGCTGGACTTCGGTGTCAACGCTGTGTCAGCAGCAAG 405
    126 TyrGluPro 128
    406 CCAGAGCGCT 414

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-642-406A-20
seq_documentation_block:

```

```

; Sequence 20, Application US/08642406A
; Patent No. 5959177
; GENERAL INFORMATION:
; APPLICANT: Hein, Mich B.
; APPLICANT: Hiatt, Andrew C.
; APPLICANT: Mg, Julian K.C.
; TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED
; SECRETORY ANTIBODIES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
; STREET: 10666 No. 595917th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642.406A
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/591,823
; FILING DATE: 02-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/427,765
; FILING DATE: 27-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Logan, April C.
; REGISTRATION NUMBER: 33,950
; REFERENCE/DOCKET NUMBER: 184.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-642-406A-20

```

```

alignment_scores:
  Quality: 166.50      Length: 103
  Ratio: 2.602        Gaps: 4
  Percent Similarity: 62.136      Percent Identity: 42.718

```

```
alignment_block:
```

```
US-09-135-238B-2 x US-08-642-406A-20 ..
```

```
Align seg 1/1 to: US-08-642-406A-20 from: 1 to: 3517
```

```

30 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 42
   ||| |||||:||||| ||| |||
238 GCGGACTGGTGCATCATCATCTACTACCCCAACCTCGTCACCGC 287
42 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
   ||| ||| :||| ||||| ||| |||
288 GCACAGCGCGGAAGTTCGGTGCAGGAA...GAGGAGAGCGCGCGTCGC 334
59 lYThrValValSerThrThrAsnPhelIleLysAlaGluTyrLysGlyArg 75
   |||:||||| |||:||||| :||| |||||:|||||
335 TCAGCGTTCGCTCG...ACGGGTACACGTCCCGAGGAATCTCCGGGAGA 381
76 ValThrLeuLysGlnTyrProArgLysAsnPhelLeuValGluValThr 92

```

```

||| :||||| |||:||||| |||:||||| |||
382 GCGAAGCTCACCAGTCCCTGATAAGGGAGTTGTGGTCACTGTTGA 431
92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 109
   |||||:||||| ||| |||:||||| |||
432 CCAGCTCACCAGCAAGCACTCAGGAGCTACAAAGTGTGGCGTGGAGTCA 481
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
   ||| ||||| |||:||||| |||:|||||
482 AC...GGCCGTGGCGTGGAGCTTCGGTGTCAACGGTGTGGTCAAGCAGGAG 528
126 TyrGluPro 128
   |||||
529 CCAGAGCGCT 537
seq_name: /cgn_2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-1
seq_documentation_block:
; Sequence 1, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 1
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; TOPOLOGY: DESCRIPTION: Rabbit polyimmunoglobulin receptor
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 124....2445
; US-08-434-000A-1

```

```

alignment_scores:
  Quality: 166.50      Length: 103

```

Ratio: 2.602 Gaps: 4
Percent Similarity: 62.136 Percent Identity: 42.718

alignment_block:
US-09-135-238B-2 x US-08-434-000A-1 ..

Align seg 1/1 to: US-08-434-000A-1 from: 1 to: 3517

30 GlyGlySerValThrIleLeuValThrPro.....GluMe 42
238 GGGAGCTGGTGTCCATACATGCTACTACCAACACCTCGTCAACCCG 287
42 tHisValArgIleTyrLeuValThrPro.....GluMe 42
288 GCACAGCGGAGTGTGGTGGCGGAA...GAGGAGAGCGCGCTGG 334
59 lyThrValValSerThrAsnPhelLeuValThrPro.....GluMe 75
335 TGACGCTGGTGTCCATACATGCTACTACCAACACCTCGTCAACCCG 381
76 ValThrLeuValThrProArgLysAsnLeuValThrPro.....GluMe 92
382 GGCAGAGCTACCGAGTTCCTCTGATAAGGGGAGTTTGTGTGCTGTGA 431
92 rGlnLeuThrGlySerAspSerGlyValThrAlaCysGlyAlaGlyMeta 109
432 CCACCTACCCAGAGAGCTACAGGAGCTACAAGTGTGGCTGGGAGTCA 481
109 snThrAspArgGlyThrGlnLysValThrLeuValThrPro.....GluMe 125
482 AC...GGCGGTGGAGTTCGCTGCTCAACGCTGTGTCAGCAGCAAG 528
126 TyrGluPro 128
529 CCAGAGCT 537

seq_name: /cgn2_5/ptodata/1/ina/5D_COMB.seq:US-08-935-450-10

seq_documentation_block:
; Sequence 10, Application US/08935450
; Patent No. 5977311
; GENERAL INFORMATION:
; APPLICANT: Nandabalan, Krishnan
; APPLICANT: Yang, Meijia
; APPLICANT: Schulz, Vincent
; TITLE OF INVENTION: 53BP2 COMPLEXES
; FILE REFERENCE: 7934-054
; CURRENT APPLICATION NUMBER: US/08/935,450
; CURRENT FILING DATE: 1997-09-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 914
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-935-450-10

alignment_scores:
Quality: 142.00 Length: 247
Ratio: 1.420 Gaps: 14
Percent Similarity: 40.486 Percent Identity: 26.721

alignment_block:
US-09-135-238B-2 x US-08-935-450-10 ..

Align seg 1/1 to: US-08-935-450-10 from: 1 to: 914

133 GlnProMetProGluThrProLysThrPheHisLeu..... 144
232 CAGCCCTCTCGGAACCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 281
145ProTyrLeu.....PheGlnMetProAlaTyrAlaSerSers 157

282 ACATCCCTGGCTCAAGCGCATACAGCCACCCACCGCTCAGCAGTAC 331
157 erLysPheValThrArgValThr.....ProAlaGlnArg..... 169
332 AGCCCTTCCACAGCGAGTTACAGCCAGCCACCTACACACCGGAGGT 381
170GlyLysValProPro.ValHisHisSerSerP 180
382 TACAGCCAGGTTACAGCGCCACCGCTCCACCTCCACCCACCACTGC 431
180 roThrGln.....lleThrHisArgProArgValSerArg 192
432 CTACAACTTGGGAGCTACGGGGTTACACCGCGGCCCTTATACCCAC 481
193 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThr 209
482 CGCCACCC...CCACCGCAGACCTACCTCAGCCCACTATAACCA 528
209 rAlaSerLysIleSerAlaLeuGlyLeuLys.....ProG 223
529 TATCAGCAGTATGCCAGCAGTGGAACTACTATCAGACCCAGGCGCAG 578
223 lInThrProSerTyrAsnHisHisThrArgLeuHisArgGlnArgAla 239
579 TGCGCCCATACTACGGGAATAC..... 601
240 AspTyrGlySerGlnSerGlyArgGluGlyGlnGlyPheHisIleLeu 256
602 GACTACGGGAGTACTCCGGGAACACACAGGGTGGCAGCAAGTACACAG 651
256 eProThrIleLeuGlyLeuPheLeuAlaLeuLeuValVal 273
652 GCCAGTG..... 658
273 ysArgAlaValGluArgArgLysAlaLeuSerArgAlaArgLeu 289
658 658
290 AlaValArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerP 306
659TGACCCAGAGGCTCCCG 675
306 oArgProArgSerGlnAsnAsnIleTyrSerAlaCys..... 318
676 GAGGCGCTCGCGGCTCTCTCCACCC...AGCGCTGCTCGGCCCTCTCT 722
319ProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAla 333
723 CTGCCCCCGCAGATCCCTGTGTCT.....GGGATGGG 757
334 ProValProGly.....ProGlyAlaProLeuProPro 344
758 GTCATCCCGAGGCTGCTCTCTCCAGCCCACTGCTCTCC 796

seq_name: /cgn2_5/ptodata/1/ina/5D_COMB.seq:US-08-824-405-5

seq_documentation_block:
; Sequence 5, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladhner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: dUTPase, its isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA


```
282 euSerArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSer 298
      ::||| :: ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2566 GAGGAGGGTGTCTCTGGAAGCAGGCTCAGCCCTCCTGGACGCA 2615
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
299 SerGln.....ArgProArgGlySe 305
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2616 TCCGGCTGTGAGTCCAGCCAGGACACAGGAGCCCGTCTGAC 2665
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
305 rProArgSerGlnAsnIleTySerAlaCysProArgArgA 322
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2666 TCCTCACCGGAGGCC.....TCTGCCCGCCCACTCATG 2700
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
322 laArgGlyAlaAsp.....AlaAlaGlyThr 330
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2701 CTAGGAGAGGCTCTCTGGCTTTTTCACACAGGCTCCGGGAGGACACA 2750
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
331 Gly..GluAlaProValProGlyPro.....GlyAlaPro 341
      ||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2751 GGCTGGATCCCTACCCAGGCGCTTCACACAGGCGCAGGTGCTGCG 2800
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
342 LeuProAlaProLeuGlnValSerGluSerProTrpLeuHisAlaPro 358
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2801 CTAGAGCTGCCAGACCAATATCCAGGAGGCCCT.....GCCCC 2841
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
358 oSerLeu 360
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2842 TGACCTA 2848

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-378-939-9

seq_documentation_block:
; Sequence 9, Application US/08378939
; Patent No. 5876961
; GENERAL INFORMATION:
; APPLICANT: CROME, JAMES SCOTT
; APPLICANT: LEWIS, ALAN PETER
; TITLE OF INVENTION: PRODUCTION OF ANTIBODIES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROTHWELL, FIGG, ERNST & KURZ
; STREET: 555 THIRTEENTH ST. N.W.
; CITY: WASHINGTON
; STATE: D. C.
; COUNTRY: U.S.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378,939
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/952640
; FILING DATE: 01-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1808-118
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 783-6040
; TELEFAX: (202) 783-6031
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
```

```
FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 35..92
FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 93..1465
FEATURE:
; NAME/KEY: CDS
; LOCATION: 35..1465
US-08-378-939-9

alignment_scores:
  Quality: 130.00      Length: 432
  Ratio: 0.660        Gaps: 22
  Percent Similarity: 45.602  Percent Identity: 22.454

alignment_block:
US-09-135-238B-2 x US-08-378-939-9 ..
Align seg 1/1 to: US-08-378-939-9 from: 1 to: 1617
1 MetAspArgTrpLeuTrpProLeuTyTrpLeuProValSerGlyAlaLe 17
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
35 ATGGAC...TGGACCTGGAGGTTCCTCTTTGTG...GTGGCAGCAGCTAC 78
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
17 uArgIleLeuProGluValLysVal.....GluGlyG 28
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
79 AGGTGTCCTCCAGTCCAGATGCAGGTGGTGCAGTCTGGGGCTGAAGTAAAGA 128
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
28 luLeuGlyGlySerValThrIleLysCysProLeuProGlu..... 41
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
129 AGCCTGGGTCTCTGGTGACGGTCTCCGTGCAAGGCATCTGGAGGCACCTTC 178
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
42 MethHisValArgIleTyTrpLeuCysArgGluMetAlaGlySerGly..... 56
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
179 AGCAACTATCTATCAGTGGGTGCGACAGGCGCCCTGGACAAGGGCTTGA 228
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
57 .....ThrCysGlyThrValValSerThrThrA 66
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
229 GTGGATGGGAGGATCATCCCTCTTTTGTGTACACCACTACTACTACAGA 278
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
66 snPheIleLysAlaGluTyLysGlyArgValThrLeuLysGlnTyPro 82
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
279 ACTTC.....CAGGGCAGAGTCACGATTACCGCGGACAAA 313
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
83 ArgLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSe 99
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
314 TCCACCAGCACAGCCACATGGAGCTGACTAGCCTGAGATCTGAGGACAC 363
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
99 rGlyValTyAlaCysGlyAlaGly.....MetAsnThrAspA 112
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
364 GCGCGTGTATTACTGTGGCAGATCGCTACAGGCGAGGCAAAATTTTGACC 413
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
112 rGlyLysThr..GlnLysValThrLeuAsnValHisSerGluTyGluPr 128
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
414 GGGCCCGGGTGGCTGGTTCGACCCCTGGGGCCAGGCGACCTGGTCACC 463
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
128 oSerTrpGluGluGlnProMetProGluThrProLysTrpPheHisLeuP 145
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
464 GTCTCCTCATCGCTCCACCAAGGCGCCATCGGTCTTCCCTCCCTGGCACCC 513
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
145 rTyTrpLeuPheGlnMetProAlaTyTrAlaSerSerSerLysPheValThr 161
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
514 CTCCAAGAGCACCTCTGGGGGCGACAGCGCCCTGGGCTGCCTGG...TCA 560
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
162 ArgValThrThrProAlaGln.....ArgGlyLysValProProVa 175
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
561 AGGACTACTTCCCGCAACCGGTGACGGTGTCTGTGGAACTCAGGCGGCCCTG 610
      |||::| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
175 lHisHisSerSerProThrThrGlnIleThrHisArgProArgValSerA 192
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
611 ACCAGCGGCGTGCACACCTTCCCGGCTGCTCTACAGTCTCTCAGGACTCTA 560
```

[illegible]

4062 CATGAACCTGGGTGAGCAGACCCATCGAGAGAACCTTGAGTGGATTGGAC 4111
70AlaGluTyrLysGly 74
4112 GTAATTAATCCTCACAATGGTGGTACTACACCAAGAGTTCAAGGAC 4161
75 ArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluVa 91
4162 AAGGCCCTTTACTGTAGACAGTCATCCAAACACAGCCTACATGGAGCT 4211
91 lThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly. 107
4212 CCTCAGTGTACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGAGGCT 4261
108MetAsnThrAspArg...GlyLysThrGlnLysValThrLeu 120
4262 ACTATTACTATTCTTTGGACTACTGGGGTCAAGGAACCTCAGTCACCGTC 4311
121 AsnValHisSerGluTyrGluProSer...TyrGluGluGlnProMetp 136
4312 TCCTCAGCTAGCACCAAGGCCCATGGTCTTCCCTCGGCACCTCCTC 4361
136 toGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMetProAla 152
4362 CAAGAGCACCTCTGGGGGCACAGCGGCCCTGG.....GCTGCC 4399
153 TyrAlaSerSerLysPheValThrArgValThrProAlaGln.. 168
4400 TGG.....TCAGGACTACTTCCCGAACCGGTG 4428
169ArgGlyLysValProProValHisHisSerSerProThrThrG 183
4429 AGCGTGTCTGTAACCTCAGCGGCCCTGACAGCGCGTGCACACCTTCCC 4478
183 InLeThrHisArgProArgValSerArgAlaSerSerValAlaGlyAsp 199
4479 GGCTGCTTACACTCTCAGACTCTACTCCCTCAGCAGCGTGTGACCG 4528
200 LysProArgThrPheLeuProSerThrAlaSerLysIleSerAlaLe 216
4529 TGCCCT.....CCAGCAGCTTGGGCACCCACACCTACATCTGC 4566
216 uGluGlyLeuLeuProGlnThrProSerTyrAsnHisHisThrArgL 233
4567 ACCTGAATCACAGCCACACCAAGGTGGACA..... 4603
233 euHisArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGly 249
4604AGAAAGTTGGTGAGAGCCAGCAC.....AGGGAGGA 4636
250 GlnGlyPheHisIleLeuIleProThrIleLeuGlyLeuPheLeuLeuAl 266
4637GGTGTCTCTGGAAGC 4653
266 aLeuLeuGlyLeuValValysArgAlaValGluArgLysAlaLeuS 283
4654 AGGCTCAGCGCTCTCGCTGGAGCC...ATCCCGGCTATGCAGCCCACT 4700
283 exArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSer 299
4701 CCAGGGCAGCAAGGCAG.....GCCCGCTCTGCC 4729
300 GlnArgProArgGlySerProArgProArgSerGlnAsnAsnIleTyrSe 316
4730 TCTTACCCGAGGACTCTGCCCGCCCACTCATGCTCAGGAGGAGGTCT 4779
316 rAlaCysProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluA 333
4780 TCTGGCTTTTCCAGGCTCTGGCAGGCACACAGCTAGGT.....G 4820
333 laProValProGlyPro.....GlyAlaProLeuProPro 344
4821 CCCTTAACCCAGGCTCTGCACACAAAGGGCAGGTGCTGGGCTCAGACCT 4870

345 AlaProLeuGlnValSerGluSerProTrpPheHisAlaProSerLeu 360
4871 GCCAAGAGCATATCCGGGAGGACCT.....GCCCTGACCTA 4909
seq_name: /cgn2_5/ptodata/1/ina/5B_COMB.seq:US-08-157-101A-6
seq_documentation_block:
; Sequence 6, Application US/08157101A
; Patent No. 5808032
; GENERAL INFORMATION:
; APPLICANT: KURIHARA, TATSUYA
; APPLICANT: MATSUKURA, SHIGEKAZU
; APPLICANT: TSURUOKA, NOBUO
; APPLICANT: ARIMA, KENJI
; APPLICANT: NISHIHARA, TATSURO
; TITLE OF INVENTION: ANTI-HBS ANTIBODY GENES AND EXPRESSION
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/157,101A
; FILING DATE: 05-APR-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: TITUS, MARLANA K
; REGISTRATION NUMBER: 35843
; REFERENCE/DOCKET NUMBER: 9437/204199
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3711
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUCH
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-157-101A-6
alignment_scores:
Quality: 125.50 Length: 432
Ratio: 0.668 Gaps: 24
Percent Similarity: 43.519 Percent Identity: 23.843
alignment_block:
US-09-135-238B-2 x US-08-157-101A-6 ..
Align seg 1/1 to: US-08-157-101A-6 from: 1 to: 1576
6 TrpProLeuTyrPheLeuProValSerGlyAlaLeuArgIleLeuProG1 22
17 TGGCCTTTTAAAGAGGTGTCAGTGTCCAGGTGCA..... 49
22 uValLysValGluGly.GluLeuGlyClySerValThrIleLysCysPro 38
50GCTGGTGGAGTCTGGGGAGGCGTGGTC..... 77
39 LeuProGluMethHisValArgIle..... 46

```
78 CAGCCTGGGAGGTCCCTGAGACTCTCTGTGCAGCGCTCTGGATTACACTT 127
47 .....TyrLeuCysArgGluMetAlaGlySerGlyThrC 58
128 CAGTAGCAATTCATGCACTGGGTCCGCCAGCGCTCCAGGCAAGGGGTGG 177
58 ysGlyThrValVal.....SerThrThrAsnPheIleLysAla 70
178 AGTGGGTGGCACTTATATATATGATGGAATCATAAATCTACGCACAC 227
71 GluTyrLysGlyArgValThrLeuLysGlnTyrProAArgLysAsnLeuPh 87
228 TCCGTGAAGGCCGATTACCACTTCCAGAGACAATTCACAAGACACACT 277
87 eLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaC 104
278 GTATCTGGNAGTGAGAGCCTGCAACTGAGGACACGGGTGTCTATTACT 327
104 ysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysVal..ThrLe 120
328 GT.....ATAAGAGATCAAACTTACGGAGTCCACAGA 359
120 uAsnValHisSerGluTyrGluProSerTrpGluGluGlnProMetPro. 136
360 TTTGACTCTCTGGGGCCAGGGAACCC...TGCTACCGTCTCTCAGCCTC 406
137 .....GluThrProLysTrpPhe.....HisLeuPro 145
407 CACCAAGGGCCCATCGGTCTCCCTCGCACCTCTCCAGAGACGACT 456
146 TyrLeuPheGlnMetProAlaTyrAlaSerSerLysPheValThrAr 162
457 CTGGGGGCACAGCGCGCTGCTGCTGCTCA.....AG 491
162 qValThrThrProAlaGln.....ArgGlyLysValProProValH 176
492 GACTACTTCCCGAACCCGGTCAGCGTGTGCGGAACCTCAGCGCGCTGGC 541
176 IsHisSerSerProThrThrGlnIleThrHisArgProArgValSerArg 192
542 CAGCGCGTGCACACCTTCCCGGTGTCTACAGTCTCAGGACTCTACT 591
193 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThrTh 209
592 CCTCAGCAGCGTGTGACCGTGCCT.....CCAGCAGCTTG 629
209 rAlaSerLysIleSerAlaLeuGluGlyLeuLysProGlnThrProS 226
630 GGCACCCAGACTACATCTGCAAGTGTAATCACAAGCCCAACACCAA 679
226 erTyrAsnHisHis.....ThrArgLeu 233
680 GTGGCAAGAAAGTTGAGCCAAATCTTGACAAACATCACACATGCC 729
234 HisArgGlnArgAlaLeu.....AspTyrGlySerGlnSerGl 246
730 CAGCGTCCCGACGACCTGAACCTCTGGGGGACCGTCACTCTCTCTCTC 779
246 yArgGluGly..GlnGlyPheHisIleLeuIleProThrIleLeuGlyLeu 262
780 CCCCCAAACCCAGGACACCTCATGATCTCCCGGACCCCTGTAGGT... 826
263 PheLeuLeuAlaLeuLeuGly..... 269
827 .....CACATGCGTGGTGGTGAGGTGAGCCACGACGACCTGAGG 867
270 .....LeuValValLysArg..... 274
868 TCAAGTTCAACTGTGTGAGCGCGGTGGAGTGCATAATGCCAAGACA 917
275 .....AlaValGluArgArgLysAlaLeuSerArgArgAla 286
918 AAGCCGGGGAGGAGCAGTACACAGCACCTACCGGGTGTGTCAGCGTCT 967
```

```
287 ArgArgLeuAla.....ValArgMetArgAl 295
968 CACGCTCTGCACCAAGGAGTGAATGGCAAGGAGTACAAGTGCAAGG 1017
295 aLeuGluSerSerGlnArgProArgGlySerProArgProArgSerGlnA 312
1018 TCTCCAAACAAGC.....CCTCCCAAGCCCCCATCGAGAA 1052
312 snAsnIleTyrSer.....AlaCysProArgArgAlaArgGlyAlaAsp 326
1053 ACCATCTCCAAAGCAAGGGCAGCCCCGAGA..... 1084
327 AlaAlaGlyThrGlyGluAlaProValProGly.....Pr 338
1085 ACCAGAGGTGTACACCTGCCCTCCCGGATGAGCTGACCAAGAACC 1134
338 oGlyAlaProLeuProAlaProLeuGlnValSerGlu 351
1135 AGTCAAGCTGACCTGCCTGGTCAAGGGCTTCTATCCAG 1174
seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-713-118-1
seq documentation block:
; Sequence 1, Application US/08713118
; Patent No. 6040436
; GENERAL INFORMATION:
; APPLICANT: Franco, Rodrigo
; APPLICANT: Sun Chen, Ai Ru
; APPLICANT: Suey, David J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING HUMAN NEURONAL
; TITLE OF INVENTION: CALCIUM CHANNEL SUBUNITS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08713,118
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mata, Elizabeth W.
; REGISTRATION NUMBER: 38,236
; REFERENCE/DOCKET NUMBER: ACC96-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7266 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 92..7102
; US-08-713-118-1
```

```
alignment_scores:
  Quality: 125.50      Length: 408
  Ratio: 0.721        Gaps: 24
  Percent Similarity: 42.647      Percent Identity: 23.775
```

```
alignment_block:
US-09-135-238B-2 x US-08-713-118-1 ..
Align seg 1/1 to: US-08-713-118-1 from: 1 to: 7266

14 SerGlyAlaLeuArgIleLeuProGluValLysValGluGlyGluLeu1 30
:::||||: ||| |||||::
5710 GCAGGCTCTCGAGGCGCTCTCCAGAT.....GG 5738
:::||||: ||| |||||::

30 yLysSerValThrIleLysCysProLeuProGluMetHisValArgIle1 47
||:||||: ||| |||||::
5739 GTCCGTGTCCTGTT.....CCACCTCTGAGGCCAC..... 5772
||:||||: ||| |||||::

47 yLeuCysArgGluMetAlaGlySerGlyThrCysGlyThrValValSer 63
||||: |||||::
5773 .....CCTGGAGCAGACAGCGGCTGTCTCCGAGG 5805

64 ThrThrAsnPhe.....IleLysAlaGluTyr1 73
:::||||: ||| |||||::
5806 AGCCGGGTTTCTTCGACAGAGTTCACCTCCCTCAGCAATGGCG 5855
:::||||: ||| |||||::

73 sGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValG 90
||||: |||||::
5856 GGGCCATACAAACCAAGAGAGTGGCATCAAGAGTCTGTCTCTGGGCG 5905
:::||||: ||| |||||::

90 Lu.....ValThrGlnLeuThrGluSerAspSerGlyVal 101
||:||||: ||| |||||::
5906 ACTCAAGACCAGCATGCCCATCAGGCCAGGCCACCCCTGGAGCG 5955
:::||||: ||| |||||::

102 Tyr.....AlaCysGlyAlaGlyMetAsnThrAspArgG1 113
:::||||: ||| |||||::
5956 TGCCCACTCCACAGAGATCCCTGTGGGCGGTGAGGACACTGGCTGTG 6005
||:||||: ||| |||||::

113 yLysThrGlnLysValThrLeuAsnValHisSerGluTyrGluProSer1 130
||:||||: ||| |||||::
6006 AGTTTCAGATGCA.....GAGCATAACCGGAGGGGCCCTGAT 6043
||:||||: ||| |||||::

130 rp.GluGluGlnProMetProGluThrProLysTrpPheHisLeuProTy 146
||:||||: ||| |||||::
6044 GGGGAGCCCGCCAGCTGGCTGGTGAGAGCCAG.....GTCG 6078
||:||||: ||| |||||::

146 rLeuPheGlnMetProAlaTyrAlaSerSerLysPheValThrArgV 163
:::||||: ||| |||||::
6079 AGCGCCTCCATGCGCGCGCTTGGCGGAGACTCAGCCCGTCAAGATG 6128
||:||||: ||| |||||::

163 alThrThrProAlaGlnArgGlyLysValProProValHisSerSer 179
||:||||: ||| |||||::
6129 CCAGC.....CCCATGAGCGCTCCATC 6151
||:||||: ||| |||||::

180 ProThrThrGlnIleThrHisArgProArgValSerArgAlaSerSerVa 196
||:||||: ||| |||||::
6152 TCCAGC.....CTGGCCCGAGCGCGCTGGGACTCATCTTTGCGAC 6195
||:||||: ||| |||||::

196 lAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLysI 213
:::||||: ||| |||||::
6196 CACCCCGGACCGCCA.....CCCCCTAGCGAGCGTGC.... 6229
||:||||: ||| |||||::

213 leSerAlaLeuGluGlyLeuLeuLysProGlnThrProSerTyrAsnHis 229
||:||||: ||| |||||::
6230 .....TCGCACCCACCAC 6241
||:||||: ||| |||||::

230 HisThrArgLeuHisArg.....GlnArgAlaLeuAspTy 241
||:||||: ||| |||||::
6242 CACCACCGTGCACCCCGCAGGAGCAGAGCAGAGTCCCTGGAGAA 6291
||:||||: ||| |||||::

241 rGlySerGlnSerGlyArgGlu..... 248
||:||||: ||| |||||::
6292 GGGGCCACAGCTGTCTCGATATGATGGCGGACCAAGCAGTGTGTGG 6341
||:||||: ||| |||||::

249 .....GlyGlnGlyPheHisIleLeuIleProThrIle 259
||:||||: ||| |||||::
6342 GGGCGGGGCTGCCCCGGGAGAGGG.....CCTACA... 6373
```

•

1

BASE COUNT	71 a	153 c	114 g	484488-4
ORIGIN				340488-343479, and

name: gb_est42:AW402953
documentation block:

WASHU-MERCK EST Project 1997
Unpublished (1997)
On Jan 14, 1998 this sequence version replaced gi:1877745.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis MO 63108

```

LOCUS      AW402953      312 bp      mRNA      EST      16-FEB-2000
DEFINITION DI-HF-BKO-06-0-UI-RI NIH_MGC_36 Homo sapiens cDNA clone
IMAGE:3055859 5', mRNA sequence.
ACCESSION  AW402953
VERSION    AW402953.1 GI:6921743
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 312)
AUTHORS   NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   On May 18, 1998 this sequence version replaced gi:3138550.
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/brp/image/image.html
Seq primer: M13 Forward.
FEATURES   Location/Qualifiers
            1..312
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3055859"
            /clone_lib="NIH_MGC_36"
            /tissue_type="lymph"
            /cell_type="germinal center B cells"
            /cell_line="MGC85"
            /lab_host="DH10B (LT1)"
            /note="vector: pT7n3-Pac; Site.1: NotI; Site.2: Eco RI;
            Constructed from size fractionated cytoplasmic mRNA
            (0.5-1.5kb). Directionally cloned. Cells provided by Louis
            M. Staudt, Ph.D. Library preparation by Maria de Fatima
            Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
            97 a 74 c 87 g 54 t
BASE COUNT 97 a 74 c 87 g 54 t
ORIGIN
alignment_scores:
    Quality: 544.00      Length: 101
    Ratio: 5.386        Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-135-238B-2 x AW402953
Align seg 1/1 to: AW402953 from: 1 to: 312
41 GluMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyTh 57
10 GAATGTCATGTGAGGATATATCTGTCCGGGAGATGCTGGATCTGGAAC 59
57 rCysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysG 74
60 ATGTGGTACCGTGGTATCCACCACTTCATCAAGGACGAATACAAGG 109
74 lYArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 90
110 GCCGAGTACTCTGACACATATCCACCACTTCATCAAGGACGAATACAAG 159
91 ValThrGlnLeuThrGluSerSerGlyValTyrAlaCysGlyAlaGlu 107
160 GTACACAGCTGCACAGAAAGTGCACGCGAGTCTATGCTCGCGAGCGGG 209
107 yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 124

```

```

210 CATGACACAGACCGGGGAAGAGACCCAGAAAGTCCCTGATGTCCACA 259
124 erGluTyrGluProSerTrpGluGlnProMetProGluThrProLys 140
260 GTGAATACGAGCCATCATGGGAAGAGACGACCAATGCTGAGACTCCAAA 309
141 Trp 141
111
310 TGG 312
seq_name: gb_est45:AW655210
seq_documentation_block:
LOCUS      AW655210      518 bp      mRNA      EST      05-APR-2000
DEFINITION 105863 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION  AW655210
VERSION    AW655210.1 GI:7421036
KEYWORDS   EST.
SOURCE     Bos taurus.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
            Bovidae; Bovinae; Bos.
REFERENCE  1 (bases 1 to 518)
AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
            Keeler,J.W.
TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
JOURNAL   Unpublished (2000)
COMMENT   On Mar 8, 1999 this sequence version replaced gi:4388119.
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904 e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 86 row: K column: 10
Seq primer: ATTAGGTGACACTATAG.
FEATURES   Location/Qualifiers
            1..518
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /clone_lib="MARC 1BOV"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /note="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
            library made from pooled tissue from lymph node, ovary,
            fat, hypothalamus, and pituitary."
BASE COUNT 124 a 133 c 157 g 104 t
ORIGIN
alignment_scores:
    Quality: 522.00      Length: 136
    Ratio: 4.424        Gaps: 0
    Percent Similarity: 86.765 Percent Identity: 68.382
alignment_block:
US-09-135-238B-2 x AW655210
Align seg 1/1 to: AW655210 from: 1 to: 518
1 MetAspArgTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLe 17
110 ATGGACCTGTGGCTTTGGCGCTTTACTTCTTCCAGTAGTCGAGGCCCC 159

```

```

17 uArlleLeuProGluValLysGluLeuGlySerValt 34
   ::::::::::::::::::::::::::::::::::::
160 GAAGGTCCTCCAGAGTAAGATGGAAGATGCTGGGAGATCCATTA 209
   ::::::::::::::::::::::::::::::::::::::::::::
34 hrileLysCysProLeuProGluMetHisValargileLysLeuValt 50
   ::::::::::::::::::::::::::::::::::::::::::::
210 CCATGAGTCCCTCTCTGAAACGATGAGGATATATCTGTGCGG 259
   ::::::::::::::::::::::::::::::::::::::::::::
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrAsnPh 67
   ::::::::::::::::::::::::::::::::::::::::::::
260 ACCATTCAGAGTCTGGAAGATGACACCGCTGTGTCCAGCAACAAATA 309
   ::::::::::::::::::::::::::::::::::::::::::::
67 eileLysAlaGluThrLysGlyArgValThrLeuLysGlnTyProArgL 84
   ::::::::::::::::::::::::::::::::::::::::::::
310 CGTCAGAGAGAAATTCAGACCGAGTCAGCGTGGAGCAGTGTCCGGACA 359
   ::::::::::::::::::::::::::::::::::::::::::::
84 yAsnLeuPheLeuValGluValThrGlnLeuThrCluserAspSerGly 100
   ::::::::::::::::::::::::::::::::::::::::::::
360 GGAATCTGTTCTGTGGTGATGACAGAGCTGACCAAGATGACAGCGGG 409
   ::::::::::::::::::::::::::::::::::::::::::::
101 ValTyAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnL 117
   ::::::::::::::::::::::::::::::::::::::::::::
410 ATCATGCTGTGGGTGGGGCGGACACAGCCGAGGCAAGACCCAGCA 459
   ::::::::::::::::::::::::::::::::::::::::::::
117 sValThrLeuAsnValHisSerGluTyGluProSerTrpGluGlnP 134
   ::::::::::::::::::::::::::::::::::::::::::::
460 GATCACCTGACTGTTACAGCGGTTTACAGAGCATCTCTGGGAAGAAGC 509
   ::::::::::::::::::::::::::::::::::::::::::::
134 roMetPro 136
   ::::::::::::::::::::::::::::::::::::::::::::
510 CGATGCT 517
   ::::::::::::::::::::::::::::::::::::::::::::

```

seq_name: gb_est8:AA521993

```

seq documentation block: 627 bp mRNA EST 17-JUL-1997
LOCUS AA521993
DEFINITION V61612.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
IMAGE:891502.5', mRNA sequence.
ACCESSION AA521993
VERSION AA521993.1 GI:2262738
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMIT Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797443.
Contact: Marra M/Mouse EST Project
WashU-HMIT Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:519462
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 494.
Location/Qualifiers
1..627
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:891502"

```

FEATURES
source

```

/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"

```

```

/note="Organ: mammary gland; Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCCGCGCAATGGTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

```

BASE COUNT 165 a 192 c 156 g 114 t
ORIGIN

alignment_scores:

Quality: 497.00 Length: 206
Ratio: 3.227 Gaps: 5
Percent Similarity: 74.757 Percent Identity: 52.913

alignment_block:

US-09-135-238B-2 x AA521993 ..

Align seg 1/1 to: AA521993 from: 1 to: 627

```

104 CysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLe 120
   ::::::::::::::::::::::::::::::::::::::::::::
10 TGTGGTGTGGCATGAACAGACAGCAAGGCAAGACCCAGAAATCACCT 59
   ::::::::::::::::::::::::::::::::::::::::::::
120 uAsnValHisSerGluTy...GluProSerTrpGluGlnProMetp 136
   ::::::::::::::::::::::::::::::::::::::::::::
60 GAATGTCCATATGAATACCCAGACCAATCTGGGAAGATGAATGAGACT 109
   ::::::::::::::::::::::::::::::::::::::::::::
136 roGluThrProLysTrpPheHisLeuProTyLeuPheGlnMetPro... 151
   ::::::::::::::::::::::::::::::::::::::::::::
110 CTGAGCGGCCAAGATGGTTCACAGATTCTCTGCAGCACAGATGCCCTGG 159
   ::::::::::::::::::::::::::::::::::::::::::::
152 .....AlaTyAlaSerSerSerLysPheValThrArgValTh 164
   ::::::::::::::::::::::::::::::::::::::::::::
160 CTCACGGGAGTGAACATCCAGCTCTCTGGAGTCATAGCCAAAGTTAC 209
   ::::::::::::::::::::::::::::::::::::::::::::
164 rThrProAlaGlnArgGlyLysValProProValHisHisSerSerPro 181
   ::::::::::::::::::::::::::::::::::::::::::::
210 CACGCCAGCTCCCAAGAGTGGAGGCCCTCCGGTTCACCGCCCTCCAGCA 259
   ::::::::::::::::::::::::::::::::::::::::::::
181 hrThrGlnLeuHisArgProArgValSerArgAlaSerSerValAla 197
   ::::::::::::::::::::::::::::::::::::::::::::
260 TCATCTCAGTACCAACATCCAGAGTTACAGAGCATTTCTGTGTGCA 309
   ::::::::::::::::::::::::::::::::::::::::::::
198 GlyAspLysProArgThrPheLeuProSerThrThrAlaSerLysIleSe 214
   ::::::::::::::::::::::::::::::::::::::::::::
310 GTACCAAGTCCCGCCGCTCTGCCAGCAACACAGCCCTCAAGAGACTC 359
   ::::::::::::::::::::::::::::::::::::::::::::
214 tAlaLeuGluGlyLeuLeuLysProGlnThrPro.....SerTyAsnH 229
   ::::::::::::::::::::::::::::::::::::::::::::
360 CACTCAGCAGCAAT.....CAGGCCCTTAGAGGCGAGCTACAGCC 399
   ::::::::::::::::::::::::::::::::::::::::::::
229 IsHisThrArgLeuHisArgGlnAlaLeuAspTyGlySerGlnSer 245
   ::::::::::::::::::::::::::::::::::::::::::::
400 ACCACACAGACTTCATGAGCAAGGACACGCCCATGGCCCACTAT 449
   ::::::::::::::::::::::::::::::::::::::::::::
246 GlyArgGluGlyGlnGly.....PheHisIleLe 255
   ::::::::::::::::::::::::::::::::::::::::::::
450 GGGAGAGACAGCCGAGGCTTCACATCCCATCCAGAAATTCACATCT 499
   ::::::::::::::::::::::::::::::::::::::::::::
255 uileProThrIleLeuGlyLeuPheLeuAlaLeuLeuGlyLeuVal 272
   ::::::::::::::::::::::::::::::::::::::::::::
500 GATTCCGACCTCTCTGGGCTTCTCTGTGTGTTCTTTGGGACTGGTG 549
   ::::::::::::::::::::::::::::::::::::::::::::

```


3334-3337, 3682-3683, 3798-3803 (IMAGE Clonoids
1323376-1323911, 1456008-1456775, 1500552-1502855);
NCI_CGAP_Kids pool 1 LHAM 3338-3342, 3722-3725, 3776-3778
(IMAGE Clonoids 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI_CGAP_Lu5 pool 1 LHAM 3575-3582,
3851-3854 (IMAGE Clonoids
1414920-1417991, 1520904-1522439); NCI_CGAP_GC4 pool 1 LHAM
3164-3167, 3716-3720, 3733-3735 (IMAGE Clonoids
1257096-1258631, 1469064-1470983, 1475592-1475743);
NCI_CGAP_Fr22 pool 1 LHAM 2457-2459, 2758-2759, 3062-3068
(IMAGE Clonoids 985608-986759, 1101192-1101959,
1217928-1220615); NCI_CGAP_Co10 pool 1 LHAM 2644-2653,
2871-2872 (IMAGE Clonoids
1057416-1061255, 1144584-1145351). (10% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (10% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (10% of the
driver population), plus a pool of 5,472 clones from
NCI_CGAP_Sub4 (IMAGE Clonoids 2723592-2728969) (70% of the
driver population). Subtraction was performed as
previously described (Bonald, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6: 791-806.
TAG_LJB=NCI_CGAP_Kids; TAG_TISSUE=kidney; TAG_SEQ=ATTC";

BASE COUNT 103 a 107 c 111 g 161 t
ORIGIN

alignment_scores:

Quality: 444.50 Length: 117
Ratio: 4.401 Gaps: 3
Percent Similarity: 86.325 Percent Identity: 80.342

alignment_block:

US-09-135-238B-2 x AW444691/rev ..

Align seg 1/1 to reverse of: AW444691 from: 1 to: 482

41 GluMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyTh 57
|||||
481 GAATCATGTGAGATATATCTGTCCGGGAGATGGCTGATCTGGAC 432
57 rCysGlyThrValValSerThrAsnPhelIleLysAlaGluTyrLys 74
|||||
431 ATGTGTACGGTGTATCCACCACCACTTCATCAAGCAGCAATACAG 382
74 IyargValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 90
|||||
381 CGCGAGTTACTCTGAAGCAATACCCAGCAAGAAATCTGCTAGTGGAG 332
91 ValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlu 107
|||||
331 GTACACAGCTGACAGAAAGTGACAGCGGAGTCTATGCTCGGAGCGGG 282
107 yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 124
|||||
281 CATGAACACAGACCGGGGAAAGCCAGAAAGTCACCTGATCTCCACA 232
124 er...GluTyrGluProSerTyrGluGluGlnProMetProGluThrPro 139
|||||
231 GTGGTGGTTCCTCCGCTGATTTGGAGGCTCAG.....ACTACCCAG 191
140 LysTyrPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSe 156
|||||
190 AAAATGTTTCTTCTGGGAAGATTG...GAACAGCCATCTAATCTGAACAG 144
156 r 156
143 T 143

seq_name: gb_est41:AW327067

seq_documentation_block:

LOCUS AW327067 359 bp mRNA EST 27-JAN-2000
DEFINITION 20516 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.
ACCESSION AW327067
VERSION AW327067.1 GI:6762988
KEYWORDS EST.
SOURCE Bos taurus.
ORGANISM

REFERENCE

AUTHORS
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 359)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
Keele,J.W.

TITLE

Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle

JOURNAL

Unpublished (2000)

COMMENT

On Jan 6, 2000 this sequence version replaced gi:6676592.
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980404.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCACGACG

Plate: 13 row: K column: 5

Seq primer: ATTAGTGACACTATAG.

FEATURES

source

1..359
Location/Qualifiers
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site:1: XbaI; Site:2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."

BASE COUNT 90 a 111 c 87 g 71 t
ORIGIN

alignment_scores:

Quality: 415.00 Length: 119
Ratio: 4.192 Gaps: 1
Percent Similarity: 83.193 Percent Identity: 64.706

alignment_block:

US-09-135-238B-2 x AW327067 ..

Align seg 1/1 to: AW327067 from: 1 to: 359

76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92
|||||
2 GTCACCTGGACGACGTGTCGGACAGGATCTGTTCTTGTGTGTGATGAC 51
92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 109
|||||
52 AGAGCTGACCAAGAAATGACAGCGGATCTATGCTGTGGGTGGCGGGA 101
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
|||||
102 ACACAGACCGAGGCAAGACCCAGCAGATCACCTGTGCTTCCAGCGTT 151
126 TyrGluProSerTyrGluGluGlnProMetProGluThrProLysTyrPp 142
|||||
152 TACGAGCCATCTCTGGGAAGAAGCGGATGCTCTGAGCTCCAGCATGTT 201

www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 983 Std Error: 0.00
 Seq primer: -40ml3 fwd. RT from Amersham
 High quality sequence stop: 473.

FEATURES

source Location/Qualifiers

1. .485
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1353883"
 /clone_lib="NCI_CGAP_GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, IgD-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
 primed with a Not I - oligo(dT) primer
 [5'-TGTTACCAATCTGAAGTGGGCGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 109 a 84 c 137 g 135 t

alignment_scores:

Quality: 305.00 Length: 59
 Ratio: 5.169 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-135-238B-2 x AA830944/rev ..

Align seg 1/1 to reverse of: AA830944 from: 1 to: 485

179 SerProThrThrGlnIleThrHisArgProArgValSerArgAlaSerSe 195
 |||||
 485 TCCCCACACCCCAATCACCACCGCCCTCGAGTGTCCAGAGCATCTTC 436
 195 rValAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerL 212
 |||||
 435 AGTAGCAGGTGACAGCCCGGAACCTCTCTGCCATCCACTACAGCCCTCAA 386
 212 ystleSerAlaLeuGluGlyLeuLysProGlnThrProSerTyrAsn 228
 |||||
 385 AAATCTCAGCTCTGGAGGGGCTGCTCAAGCCCGCAGAGCCGACGCTACAAC 336
 229 HisHisThrArgLeuHisArgGlnArg 237
 |||||
 335 CACACACCCAGGCTGCACAGGCAGAGG 309

seq_name: gb_est8:AA491123

seq_documentation_block: 179 bp mRNA EST 18-AUG-1997
 LOCUS AA491123
 DEFINITION aa46904.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:824022 5',
 mRNA sequence.

ACCESSION AA491123
 VERSION AA491123.1 GI:2220296

KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 179)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index
 JOURNAL Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -28ml3 rev1 RT from Amersham

High quality sequence stop: 116.

FEATURES

Location/Qualifiers

1. .179

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:824022"

/clone_lib="NCI_CGAP_GCB1"

/tissue_type="germinal center B cell"

/lab_host="DH10B"

/note="vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, IgD-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
 primed with a Not I - oligo(dT) primer
 [5'-TGTTACCAATCTGAAGTGGGCGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 46 a 69 c 34 g 30 t

ORIGIN

alignment_scores:

Quality: 299.00 Length: 59
 Ratio: 5.068 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-135-238B-2 x AA491123 ..

Align seg 1/1 to: AA491123 from: 1 to: 179

163 ValThrThrProAlaGlnArgGlyLysValProProValHisHisSerSe 179
 |||||
 3 GTTACCACACAGCATCAAGGGGCAAGGTCCCTCCAGTTCACCATCTCTC 52
 179 rProThrThrGlnIleThrHisArgProArgValSerArgAlaSerSe 196
 |||||
 53 CCCCACCCCAATCACCACCGACCTCGAGTGTCCAGAGCATCTTCAG 102
 196 alaAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLys 212
 |||||
 103 TAGCAGGTGACAAAGCCGCAACCTCTCTGCCATCCACTACAGCTCAAAA 152
 213 IleSerAlaLeuGluGlyLeuLys 221
 |||||
 153 ATCTCAGCTCTGGAGGGGCTGCTCAAG 179

seq_name: gb_est40:AW291041

seq_documentation_block:

LOCUS AW291041 561 bp mRNA EST 16-JAN-2000
 DEFINITION UI-H-B12-ag1-h-06-0-UI.s1 NCI_CGAP_Sub4 Homo sapiens cDNA clone
 IMAGE:2724707 3', mRNA sequence.

ACCESSION AW291041

VERSION AW291041.1 GI:6697677

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 561)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1998 this sequence version replaced gi:3137211.
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
The sequence contained an oligo-dn track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LUNL at:
www.bio.lnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.

FEATURES source
Location/Qualifiers
1..561
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2724707"
/clone_lib="NCI_CGAP_Sub4"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pTT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; The
NCI_CGAP_Sub4 library is a subtracted library derived from
the NCI_CGAP_Sub2 library which is a subtracted library
derived from the NCI_CGAP_Sub1 library, which is a
subtracted library derived from BI. Bi constitutes a
mixture of 21 normalized or subtracted NCI_CGAP
libraries: NCI_CGAP_Co4, NCI_CGAP_Pr22, NCI_CGAP_Pr28,
NCI_CGAP_Co10, NCI_CGAP_Co16, NCI_CGAP_Kid5,
NCI_CGAP_Kid12, NCI_CGAP_Kid3, NCI_CGAP_Kid11,
NCI_CGAP_Lym2, NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_CLL1,
NCI_CGAP_Le12, NCI_CGAP_Brn23, NCI_CGAP_Lu5,
NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
NCI_CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI_CGAP_Kid3 pool 1 : LLAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911,
1456008-1456775, 1500552-1502855) NCI_CGAP_Kid5 pool 1 :
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1323931, 1471368-1472903, 149104-1493255)
NCI_CGAP_Lu5 pool 1 : LLAM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439) NCI_CGAP_GC4
pool 1 : LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonoids 1257096-1258631, 1469064-1470983,
1475592-1476743) NCI_CGAP_Pr22 pool 1 : LLAM 2457-2459,
2758-2759, 3062-3088 (IMAGE Clonoids 985608-986759,
1101192-1101959, 1217928-1220615) NCI_CGAP_Co10 pool 1 :
LLAM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255,
1144584-1145351) Subtraction was performed as previously
described [Bonaldi, Lennon & Soares (1996): Normalization
and Subtraction: Two Approaches To Facilitate Gene
Discovery. Genome Research 6, 791-806.];
TAG_LIB=NCI_CGAP_Co10; TAG_TISSUE=colon; TAG_SEQ=AAACG"
121 a 125 c 120 g 194 t 1 others

alignment_scores:
Quality: 292.50 Length: 91
Ratio: 3.953 Gaps: 3
Percent Similarity: 81.319 Percent Identity: 72.527

alignment_block:
US-09-135-238B-2 x AW291041/rev ..
Align seg 1/1 to reverse of: AW291041 from: 1 to: 561
67 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 83
|||||
561 TTCATCAGGCGAGATACAGGCCCGAGTTACTCTGAAGCAATACCCACG 512
|||||
83 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 100
|||||
511 CAAGAATCTGTTCTAGTGGAGGTAACACAGCTGACAGAAAAGTACACAGC 462
|||||
100 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 116
|||||
461 GAGTCTATGCTCGGAGCGGGCGATGAACACAGACCGGNGAAGACCCAG 412
|||||
117 LysValThrLeuAsnValHisSer...GluTyrGluProSerTTPGluG1 132
|||||
411 AAAGTCACCCCTGAATGTCCACAGTGGTAGGTTCTCCGCTGATTGGAGGC 362
|||||
132 uGlnProMetProGluThrProLysTyrPheHisLeuProTyrLeuPheG 149
|||||
361 TCAG.....ACTACCCAGAAAAATGTTTCACCTGGGAAGATTG...G 324
|||||
149 InMetProAlaTyrAlaSerSer 156
|||||
323 AACAGCCATCTAACTGAACAGT 301

OM of: US-09-135-238b-2 to: Issued_Patents_NA:* out_format : pfs
Date: Sep 12, 2000 6:53 AM
About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPFO.spool/US09135238/runat_24082000_091324_17895/app_query.fasta_1.1282
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosum62 -TRANS=human40 cdi -LIST=45 -DOCALLIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=100000
-USER=US09135238_cgn1_1_68 -NCP0=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-135-238b-2
Query length: 390
Database: Issued_Patents_NA.*
Database Sequences: 243080
Database length: 68777915
Search time (sec): 93.130000

score list:

Sequence	Strd Orig	ZScore	EScore	Len
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-7	1	198.50	262.73	3085
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-9	1	185.50	243.58	3269
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-3	1	182.00	239.68	2919
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-5	1	182.00	237.20	3630
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-642-406A-21	1	166.50	221.76	1875
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-642-406A-20	1	166.50	221.76	7.3e-05
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-6	1	166.50	214.59	0.0002
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-1	1	142.00	193.65	0.0037
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-824-405-5	1	136.50	178.54	0.0187
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-27	1	135.00	155.12	0.3776
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-32	1	131.00	147.85	0.9588
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-378-939-9	1	130.00	169.36	0.0608
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-18	1	129.00	146.21	1.18
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-157-101A-6	1	125.50	162.98	0.1377
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-713-118-1	1	125.50	145.55	1.29
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-286-740-3	1	124.50	145.24	1.34
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-957-09576-3	1	124.50	145.24	1.34
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-576-628A-2	1	123.00	140.67	2.41
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-476-273-2	1	123.00	139.14	2.93
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-476-273-2	1	123.00	139.14	2.93
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-478-967A-3	1	123.00	139.14	2.93
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-819-866-2	1	123.00	130.88	8.45
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-023-715-2	1	123.00	130.88	8.45
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-480-036-1	1	122.50	154.29	0.4200
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-461-968A-1	1	122.50	154.29	0.4200
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-462-571-1	1	122.50	154.29	0.4200
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-462-571-1	1	122.50	154.29	0.4200
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-916-917-1	1	122.50	146.09	1.20
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-972-631-1	1	122.50	146.09	1.20
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-972-631-1	1	122.50	146.09	1.20
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-972-630-1	1	122.50	146.09	1.20
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-672-211-1	1	122.50	146.09	1.20
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-09-225-170-1	1	122.50	146.09	1.20
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-455-543A-7	1	121.50	139.47	2.81
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-193-078B-7	1	121.50	139.47	2.81
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-223-305C-7	1	121.50	139.47	2.81
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-149-097D-7	1	121.50	139.47	2.81
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-310-370-1	1	121.00	159.52	0.2147
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-955-937A-1	1	119.50	149.56	0.7705
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-41	1	119.50	132.09	7.24

/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-764-233A-1 - 118.50 113.31 80.51 49377
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-455-543A-8 + 118.00 134.57 5.26 7175
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-193-078B-8 + 118.00 134.57 5.26 7175
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-223-305C-8 + 118.00 134.57 5.26 7175
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-149-097D-8 + 118.00 134.57 5.26 7175

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-7

seq_documentation_block:
: Sequence 7, Application US/08434000A
: Patent No. 6046037
: GENERAL INFORMATION:
: APPLICANT: ANDREW C. HIATT, JULIAN
: APPLICANT: K.-C. MA, THOMAS LEHNER
: TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
: TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: 633 West Fifth Street
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90071
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: MEDIUM TYPE: storage
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: IBM P.C. DOS 5.0
: SOFTWARE: Word Perfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/434,000A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA: including application
: PRIOR APPLICATION DATA: described below:
: APPLICATION NUMBER: 08/367,395
: FILING DATE: 12/30/94
: ATTORNEY/AGENT INFORMATION:
: NAME: Guise, Jeffrey W.
: REGISTRATION NUMBER: 34,613
: REFERENCE/DOCKET NUMBER: 212/127
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 552-8400
: TELEFAX: (619) 552-0159
: TELEX: 67-3510
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3095 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: TOPOLOGY: DESCRIPTION: Mouse Polyimmunoglobulin Receptor
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 85.....2400
US-08-434-000A-7

alignment_scores:
Quality: 198.00 Length: 168
Ratio: 1.980 Gaps: 5
Percent Similarity: 59.524 Percent Identity: 31.548

alignment_block:

US-09-135-238B-2 x US-08-434-000A-7 ..

Align seg 1/1 to: US-08-434-000A-7 from: 1 to: 3095

10 PheLeuProValSerGlyAlaLeuArgIleLeuProGluValLysValG1 26

```

121 TTTTCAGGGGCTCTCCAAAAAGCCCATATTTGGTCCCGAGAGGTGAG 170
26 uGlyGluLeuGlyGlySerValThrIleLysCysProLeuProGlu... 41
171 TAGTATAGAGGGGAGCTCTGTTCCATCAGCTGCTACTACCCAGACACT 220
42 .....MetHisValArgIleTyrLeuCysArgGluMetAlaGlySer 55
221 CTGTACAGCGCACCCGGAATAGTGTGCCGACAGAGGCC...AGC 267
56 GlyThrCysGlyThrValValSerThrAsnPhelIleLysAlaGluTyr 72
268 GCGATGTCACAGGCTCATCTCTTCAATGGCTACCTCTCCAAAGGAGTA 317
72 rLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuV 89
318 TTCAGCAGAGCCAACTCATCACTTCCCGAGAGAACACACATTGTGA 367
89 aGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGly 105
368 TPACATATGACACCTCACCCAGAGACACTGGAGCTACAAAGTGTGC 417
106 AlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVa 122
418 CTGGGT...ACCAGTAACCGAGGCGCTCTCTTCGATGTACAGCTGGAGST 464
122 lHisSerGluTyrGluProSerTrpGluGluGlnProMetProGluThrP 139
465 C.....AGCCAGGTTCCTCAGTGC 484
139 rLysTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSer 155
485 CGAGTGACACCCAGCTC.....TACACAAAG 510
156 SerSerLysPheValThrArgValThrProAlaGlnArgGlyLysVa 172
511 GACATAGCAGAAATGTGACCAATTGAATGCCCTTTCAAAAGGGAGATGT 560
172 lPro 173
561 TCC 564

seq_name: /cgn2_6/ptodata/L/ina/6_COMB.seq:us-08-434-000A-9
seq_documentation_block:
; Sequence 9, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

```

; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 1
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 74...2383
; US-08-434-000A-9

alignment_scores:
Quality: 185.50 Length: 164
Ratio: 1.855 Gaps: 6
Percent Similarity: 60.976 Percent Identity: 30.488

alignment_block:
US-09-135-238B-2 x US-08-434-000A-9 ..
Align seg 1/1 to: US-08-434-000A-9 from: 1 to: 3269
30 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 42
170 GGTAACTCGGTCTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACGG 219
42 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
220 GCACACCGGAAATACTGTGTCGCGACAGAGGCC...AAGGCTACTCGG 266
59 lYThrValValSerThrAsnPhelIleLysAlaGluTyrLysGlyArg 75
267 CAACCCCTCATCTCTCAAAATGGCTACTCTCGAAGGAGTATTGAGGAGA 316
76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92
317 GCCAGCCTCATCACTTCCAGAGAAATAGCACATTGTGATTAAACATTGC 366
92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 109
367 ACACTCTACCCAGAGGAGACTGGGAGCTACAAGTGTGCTCTGGGT...A 413
109 sThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
414 CCACCTAACCGAGGCGCTGTTTTCGATGTACGCTCGAGGAGTTC... 454
126 TyrGluProSerTrpGluGluGlnProMetProGluThrProLysTrpPh 142
455 .....AGCCAGGTTCCTGAGTTCCTCAAAATGACAC 483
142 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 159
484 CCATGTC.....TACACAAAGG 500
159 heValThrArgValThrThrProAlaGlnArgGlyLysValProVal 175
501 ACATAGCAGCAACTGTGACCAATGCGCTTTCAAAAGAGGGAATGCT 550
176 HisHisSerSerProThrThrGlnIleThrHisArgPro 188

```



```
; Sequence 20, Application US/08642406A
; Patent No. 5959177
; GENERAL INFORMATION:
; APPLICANT: Hein, Mich B.
; APPLICANT: Hiatt, Andrew C.
; APPLICANT: Ma, Julian K.C.
; TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED
; TITLE OF INVENTION: SECRETORY ANTIBODIES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
; STREET: 10666 NO. 5959177th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,406A
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/591,823
; FILING DATE: 02-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/427,765
; FILING DATE: 27-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Logan, April C.
; REGISTRATION NUMBER: 33,950
; REFERENCE/DOCKET NUMBER: 184.2
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-642-406A-20
```

```
alignment_scores:
  Quality: 166.50      Length: 103
  Ratio: 2.602         Gaps: 4
  Percent Similarity: 62.136  Percent Identity: 42.718
```

```
alignment_block:
US-09-135-238B-2 x US-08-642-406A-20
```

Align seg 1/1 to: US-08-642-406A-20 from: 1 to: 3517

```
30 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 42
   ||| |||||:||||| ||| |||
238 GCGAGCTCGGTGCTCATCATCTACTACCCACACCTCGTCACCCG 287a
   ||| |||||:||||| ||| |||
42 thisValArgIleLysCysArgGluMetAlaGlySerGlyThrCysG 59
   ||| ||| :||| ||||| ||||| |||
288 GCACAGCCGCGAAGTCTCGTCCGGGAA...GAGGAGAGCGCGCTGCG 334
   |||:|||||:||||| |||||:|||||:|||||:|||||
59 IyThrValValSerThrThrAsnPhelIleLysAlaGluThrLysGlyArg 75
   |||:|||||:|||||:|||||:|||||:|||||:|||||
335 TGACCGCTTGCCTCG...ACCGGCTACAGTCCCGAGGAATACCGGGAGA 381
   |||:|||||:|||||:|||||:|||||:|||||:|||||
76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92
```

```
||| :|||||:||||| |||||:||||| |||
382 GGCAGACTCACCAGCTCCCTGTATAAGGGGAGTTTGTGCTGACTGTGA 431
   |||:|||||:|||||:||||| ||| |||||:|||||:|||||
92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 109
   |||||:|||||:|||||:||||| ||| |||||:|||||:|||||
432 CCAACTCACCCAGACAGACTCAGGGAGCTACAAGTGTGCGTGGGAGTCA 481
   ||| |||||:|||||:||||| ||| |||||:|||||:|||||
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisserGlu 125
   || ||||| |||||:|||||:||||| ||| |||||:|||||:|||||
482 AC...GCCGTGGCGTGGAGTTCGGTGTCAACGTGCTGTGTCAGCCAGAAG 528
   ||||| |||||
126 TyrGluPro 128
   |||||
529 CCAGAGCCT 537

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-1

seq_documentation_block:
; Sequence 1, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; CITY: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 1
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 124.....2445
; US-08-434-000A-1
```

```
alignment_scores:
  Quality: 166.50      Length: 103
```

Ratio: 2.602 Gaps: 4
Percent Similarity: 62.136 Percent Identity: 42.718

alignment_block:

US-09-135-238B-2 x US-08-434-000A-1 ..

Align seg 1/1 to: US-08-434-000A-1 from: 1 to: 3517

```
30 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 42
   |||||:||||| |||
238 GCGCACTCGGTGCCATCATCATCTACTACCAACAACCTCCGTCACCG 287
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
42 ThisValArgIleThrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
288 GCACGCGCGGAAGTTCTGTGTCGCCGGAA...GAGGAGAGCGCGCTCG 334
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
59 lYThrValValSerThrAsnPheIleLysAlaGluThrLysGlyArg 75
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
335 TGACGCTTGCTCG...ACGCGCTACACGTCGCCAGGATACTCCGGGAGA 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 GCGAAGCTCACCGACTCCCTGATAAAGGGAGTTGTGTGAGTGTGA 431
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 109
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
432 CCAACTCACCAAGACGACTCAGGAGCTACAAGTGTGGCGTGGAGTCA 481
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
482 AC...GGCCGTGCGTGGAGTTCGGTGTCAAGTGTGTCAGCCAGAAG 528
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
126 TyrGluPro 128
   |||||
529 CCAGAGCGCT 537
```

seq_name: /cgn2_5/ptodata/1/ina/5D_COMB.seq:US-08-935-450-10

seq_documentation_block:

```
; Sequence 10, Application US/08935450
; Patent No. 5977311
; GENERAL INFORMATION:
; APPLICANT: Nandabalan, Krishnan
; APPLICANT: Yang, Meijia
; APPLICANT: Schuiz, Vincent
; TITLE OF INVENTION: 53BP2 COMPLEXES
; FILE REFERENCE: 7934-054
; CURRENT APPLICATION NUMBER: US/08/935,450
; CURRENT FILING DATE: 1997-09-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 914
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-935-450-10
```

alignment_scores:
Quality: 142.00 Length: 247
Ratio: 1.420 Gaps: 14
Percent Similarity: 40.486 Percent Identity: 26.721

alignment_block:

US-09-135-238B-2 x US-08-935-450-10 ..

Align seg 1/1 to: US-08-935-450-10 from: 1 to: 914

```
133 GlnProMetProGluThrProLysTrpPheHisLeu..... 144
   ||||| ||| |||||
232 CAGCCCTGCTCGGAACCCCGCCAGGCGCAGACCTACAATAAGAACAGCA 281
   ||||| ||||| ||||| |||||
145 ....ProTyrLeu.....PheGlnMetProAlaTyrAlaSerSers 157
```

```
|||||:||||| :||| ||| :||| :
282 ACATCCCTGGCTCAAGCCCAATACCAAGCCCGCCACCGCTCAGCAGTAC 331
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
157 erLysPheValThrArgValThr.....ProAlaGlnArg..... 169
   :||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
332 AGCCCTTCACAGCCGAGTTACAGCCAGCCACCTACACACAGGGAGGT 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
170 .....GlyLysValProPro.ValHisHisSerSerP 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 TACAGCCAGGTTTACAGAGCCCGCTCCACCTCCACCTCCACCTGC 431
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
180 rofThrGln.....IleThrHisArgProArgValSerArg 192
   |||||:||||| |||||:||||| |||||:||||| |||||:|||||
432 CTACAACTATGGAGCTACGCGGTACACCCGGCCCCCTATACCCAC 481
   |||||:||||| |||||:||||| |||||:||||| |||||:|||||
193 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThrTh 209
   :||||| :||||| :||||| :||||| :
482 CGCCAGCCC...CCAGCGCACAGACTACCTCAGCCCAACTATAACCG 528
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
209 rAlaSerLysIleSerAlaLeuGluGlyLeuLys.....ProG 223
   :|||||:||||| |||||:||||| |||||:||||| |||||:|||||
529 TATCAGCAGTATGCCAGCTGGACCACTACTATCAGACCAAGCCGCG 578
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
223 lnThrProSerTyrAsnHisThrArgLeuHisArgGlnArgAlaLeu 239
   :||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
579 TGGCGCATACTACGGGAATAC..... 601
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
240 AspTyrGlySerGlnSerGlyArgGluGlyGlnGlyPheHisIleLeuI 256
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
602 GACTACGGAGGTACTCCGGGAACACACAGGGTGGCACAAGTACACAGTA 651
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
256 eProThrIleLeuGlyLeuPheLeuAlaLeuLeuGlyLeuValValL 273
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
652 GCCAGTG..... 658
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
273 ysArgAlaValGluArgLysAlaLeuSerArgArgAlaArgLeu 289
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
658 ..... 658
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
290 AlaValArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerPr 306
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
659 .....TGACCCAGAGGCTCCCG 675
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
306 oArgProArgSerGlnAsnIleTyrSerAlaCys..... 318
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
676 GAGGCCCTGCGCGCTTCCTCCACC...AGCGCTGCTCGGCCCTCCT 722
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
319 .....ProArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAla 333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
723 CTGCCCGCCAGATCCCGTGTCT.....GGGATGGG 757
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
334 ProValProGly.....ProGlyAlaProLeuProPro 344
   :||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
758 GTATCCAGGCGCTCCCTCCCTCCAGCCACTGCTCCC 796
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-824-405-5

seq_documentation_block:

```
; Sequence 5, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: Duprase, Its Isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
```

```

; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824.405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1683 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-824-405-5

alignment_scores:
  Quality: 136.50      Length: 265
  Ratio: 1.034        Gaps: 15
  Percent similarity: 49.811  Percent Identity: 25.660

alignment_block:
US-09-135-238B-2 x US-08-824-405-5/rev ..
Align seg 1/1 to reverse of: US-08-824-405-5 from: 1 to: 1683
162 ArgValThrThrProAlaGln.....ArgGlyLysValProPro...Va 175
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1524 CGGAGCGCGCGCCCTCCAGCCCTCCCGCGCGGAGTCCCGCTCACCT 1475
175 lHisHisSerSerProThrThrGlnlleThrHisArgProArgValSerA 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1474 GTACAGTCTGTAGCCCGCGCGCGCG.....GAGCCCGCGGTGGGG 1431
192 rgAlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThr 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1430 CCGTGGGTGCTCGAGAGCGCGGCAAGCGGAGCTGCATGCCGCCACC 1381
209 ThrAlaSerLys.....lleSerAlaLe 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1380 TCCGAGCGCGCGCGCGCTTACTGGTGAATGGCGGTGTCTTTCAGA 1331
216 uGlyLeuLeuLysProGlnThrProSerTyrAsnHis..... 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1330 GCAGGGCATGCAGCAGAGCAAGACGCGAGGAGACCAACCGGGGAAGA 1281
231 .....ThrArgLeuHisArgGlnArgAlaLeuAspTyrGly 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1280 GGAGGCCACGATGAGGTGCGCGGAGGAGCGGCCCAAAACCGAAATT 1231
243 SerGln..... 244
|||||:
1230 TCCGCCAGCCCGCCGCCACCACCGCCCTCCCGGGGAGCAGTCCACAG 1181
245 SerClyArgGluGlyGlnGlyPheHislleLeuIleProThrIleLeuG 261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1180 TTCTGGTGAGAGGAGCGCGGATTAC.....ATTGTG 1146
261 lyLeu.....PheLeuLeuAlaLeuGlyLeu..... 270

```

```

|||||  |||  ...  |||:|||||:
1145 GTTTGAAGCCCTTCTCGAAAGCAGAGTTTGGGATGGGAAACTACTGGGTC 1096
271 .....ValValLysArgAlaValGluArgArgLysAl 281
|||||:|||||:|||||:|||||:
1095 CCAGTCAACTTTCACATTCAGCGCGTGGGCCAACATTTGAAGGGGAGCT 1046
281 aLeuSerArgArgAlaAaArgLeuAlaValArgMetArgAlaLeuGln 298
|||||:|||||:|||||:|||||:
1045 TCTAGGGCGTGC.....CTCACACTAGAAAGCCGCCCCCT 1008
298 erSerGlnArgProArgLysProArg.....ProArgSerGln 311
|||||:|||||:|||||:|||||:
1007 CGCCAGGGGCCCTAATGACCAACCCACCCGCCCGCCCGCGCGCAG 958
312 AsnAsn.....lleTyrSerAlaCysProArgArgAlaArgG1 324
|||||:|||||:|||||:|||||:
957 AACCAACGCTCGCCACCCCTACTTCCCAACAGCTCAAGCGCGGTGATTC 908
324 yAlaAspAlaAlaGlyThrGlyGluAlaPro...ValProGlyProGlyA 340
|||||:|||||:|||||:|||||:
907 CTTCCAGAGCGCGGAGCCCTCCCGCCCTTCTCTACCGCGCCCGCGG 858
340 laProLeuProProAla.....ProLeuGln..... 348
|||||:|||||:|||||:|||||:
857 CGGGCTTCCCGCCCTTAGGAAGCTCGCCCTTCCAGCAAGCGCGCCCG 808
349 .....ValSerGluSerProThrPheHisAlaProSerLeu 360
|||||:|||||:|||||:|||||:
807 ACTGTACTGCTCGCGGCGAGCTTGGCTCAGCGCGGCGGCGCTG 763

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-27
seq_documentation_block:
; Sequence 27, Application US/08444644
; Patent No. 601555
; GENERAL INFORMATION:
; APPLICANT: Friden, Phillip M.
; TITLE OF INVENTION: TRANSFERRIN RECEPTOR SPECIFIC
; TITLE OF INVENTION: ANTIBODY-NEUROPHARMACEUTICAL OR DIAGNOSTIC AGENT
; NUMBER OF INVENTIONS: 46
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,644
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/232,246
; FILING DATE: 07-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/800,458
; FILING DATE: 26-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US90/05077
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/404,089
; FILING DATE: 07-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Wagner, Richard W.
; REGISTRATION NUMBER: 34,480

```

REFERENCE/DOCKET NUMBER: ALK88-15AAAZ

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 10785 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

IMMEDIATE SOURCE:

CLONE: PAH4625

FEATURE:

NAME/KEY: misc.feature

LOCATION: 1..10785

OTHER INFORMATION: /note= "Function = "Expression

OTHER INFORMATION: Vector Coding Sequence"

US-08-444-644-27

alignment_scores:

Quality: 135.00 Length: 446
Ratio: 0.763 Gaps: 23
Percent Similarity: 39.686 Percent Identity: 22.646

alignment_block:

US-09-135-238B-2 x US-08-444-644-27 ..

Align seg 1/1 to: US-08-444-644-27 from: 1 to: 10785

4 TrpLeuProMetProGluThrPheLeuProValSerGlyAlaLeuArgIleLeu 20 *

1854 TGAGCTGGTGAATGCTCTTCCTC...CTGCAGGAAGTGCAGGTGGCG 1900

20 uProGluValLysValGlu.....GlyGluLeu.....GlyG 31

1901 CTCTGAGGTCCAGCTGCAACAGCTCTGGACCTGAACCTGGTGAAGCTGGAG 1950

31 LySerValThrIleLysCysProLeuProGluMetHisValArgIleLys 47

1951 CTTCATGAGAGTTCCTGC..... 1970

48 LeuCysArgGluMetAlaGlySerGlyThrCysGlyThrValValSerTh 64

1971AAGGCTTCTGTTACTCATTCTGCTGC.....TACAC 2002

64 rThrAsnPheIleLys..... 69

2003 CATGAACCTGGGTGAAGCAGAGCCATGGAGAACCTTGAGTGGATTGGAC 2052

70AlaGluThrLysGly 74

2053 GTATTAACTCTCAATGTTGGTGTACTACACAGCAAGTTCAAGAC 2102

75 ArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluVa 91

2103 AAGGCCCTTAACTGTAGCAAGTCAATCCACACAGCCCTACATAGAGCT 2152

91 lThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly. 107

2153 CCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGAGGCT 2202

108MetAsnThrAspArg...GlyLysThrGlnLysValThrLeu 120

2203 ACTATTACTATCTTTGGACTACTGGGTCAAGGAACCTCAGTCAACGCTC 2252

121 AsnValHisSerGluTyrGluProSer..... 129

2253 TCTCAGTACAGCAAGGCCCATCGGTCTTCCCTTGGCGCCCTGCTC 2302

130TrpGluG 132

2303 CAGGAGCACCCTCGAGAGCAGAGCGGCCCTGGGCTGCCTGTGTCAGGACT 2352

132 luGlnProMetProGluThrProLysTrpPheHisLeuProTyrIleuPhe 148

2353 ACTTCCCGAACCGGTGCGGTGTGGTGA...CTCAGGCGCTCTGAC 2398

149 GlnMetProAlaTyrAlaSerSerSerLysPheValThrArgValThrTh 165

2399 CAGCGCGTGCACACCTTCCAGCTGTCCAGTGTCTACAGTCTCCTCAGGACTTACT 2448

165 rProAlaGlnArgGly.LysValProProValHisHisSerSerPro... 180

2449 CCTCAGCAGCGTGTGACCGTGCCTCCAGCAACTTCGGCACCCAGACC 2498

181 ...ThrThrGlnIleThrHisArgProArgValSerArgAlaSerSerVa 196

2499 TACACCTGCACAGTAGATCACAGCCCAACACCAAGGTGGACAAGAC 2548

196 lAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLysI 213

2549 AGTTGGTGAGAGGCCA.....GCTCAGGGAGGAGGG 2580

213 lSerAlaLeuGluGlyLeuLysProGlnThrProSerTyrAsnHis 229

2581 TGTCTGCT..... 2588

230 HistThrArgLeuHisArgGlnArgAlaLeuAspTyrGlySerGlnSerGl 246

2589GGAAGCCAGGCTCA 2602

246 yArgGluGlyGlnGlyPheHisIleLeuIleProThrIleLeuGlyLeuP 263

2603 G.....CCCTCCTGCTGCAGC.... 2618

263 heLeuLeuAlaLeuLeuGlyLeuValValLysArgAlaValGluArgArg 279

2618 2618

280 LysAlaLeuSerArgArgAlaArgAlaValArgMetArgAlaLe 296

2619GCACCCCGCTGTGCAGCC 2638

296 uGluSerSerGln.....ArgProArgGlySerProArgProArgSerG 311

2639 CAGCCAGGCGCAGCAGGAGCCCATCTGCTCTCTCACCAGGAGGCC. 2687

311 lAsnAsnIleTyrSerAlaCysProArgArgAlaArgGlyAlaAsp... 326

2688TCTGCCGCCCTCATCTGCTCAGGAGAGGGTCTT 2723

327AlaAlaGlyThrGly..GluAlaProVal 335

2724 CTGGCTTTTCCAGCAGGCTCCAGGAGGAGCAGGCTGGTGGCCCTTACC 2773

336 ProGlyPro.....GlyAlaProLeuProProAlaProLe 347

2774 CCAGGCGCTTCACACAGAGGCGAGGTGTGCTGCTCAGACCTGCCAAGAAG 2823

347 uGlnValSerGluSerProTrpLeuHisAlaProSerLeu..... 360

2824 CCATATCCGGGAGGACCT.....GCCCTGACCTTAGCCGACCC 2864

361 ..LysThrSerCysGluTyrValSerLeuTyrHis 371

2865 AAAGGCCAAACTGTCCACTCCCTCAGCTGGGACAC 2899

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-32

seq_documentation_block:

; Sequence 32, Application US/08444644

; Patent No. 6015555

; GENERAL INFORMATION:

; APPLICANT: Friden, Phillip M.


```
282 euSerArgArgAlaArgLeuAlaValArgMetArgAlaLeuGluSer 298
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2566 GAGGAGGGGTGCTGCTGGAAGCAGGCTCAGCCCTCGCTGGACGCA 2615
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
299 SerGln.....ArgProArgGlySe 305
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2616 TCCGGGCTGTCAGTCCAGCCAGGCGCAGGCGCCGCTGTGAC 2665
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
305 rProArgProArgSerGlnAsnAsnIleTyrSerAlaCysProArgArgA 322
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2666 TCCTCACCGGAGGC.....TCTGCCGCGCCCACTCATG 2700

322 laArgGlyAlaAsp.....AlaAlaGlyThr 330
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2701 CTCAGGAGAGGGTCTTCTGGCTTTTTCACAGGCTCCGGCGAGGACA 2750
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
331 Gly..GluAlaProValProGlyPro.....GlyAlaPro 341
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2751 GGCTGGATGCCCTACCCAGGCCCTTCACACAGGGCGAGGTGCTGG 2800
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
342 leuProProAlaProLeuGlnValSerGluSerProTrpLeuHisAlaPr 358
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2801 CTCAGAGTGCACAGAGCCATATCCAGGAGGACCCT.....GCCCC 2841

358 oSerLeu 360
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2842 TGACCTA 2848
```

seq_name: /cn2_6/ptodata/1/ina/5C_COMB.seq:US-08-378-939-9

```
seq_documentation_block:
; Sequence 9, Application US/08378939
; Patent No. 5876961
; GENERAL INFORMATION:
; APPLICANT: CROWE, JAMES SCOTT
; APPLICANT: LEWIS, ALAN PETER
; TITLE OF INVENTION: PRODUCTION OF ANTIBODIES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROTHWELL, FIGG, ERNST & KURZ
; STREET: 555 THIRTEENTH ST. N.W.
; CITY: WASHINGTON
; STATE: D. C.
; COUNTRY: U.S.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378,939
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/952640
; FILING DATE: 01-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1808-118
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 783-6040
; TELEFAX: (202) 783-6031
; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
```

```
;
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 35..92
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 93..1465
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 35..1465
; US-08-378-939-9

alignment_scores:
  Quality: 130.00      Length: 432
  Ratio: 0.660        Gaps: 22
  Percent Similarity: 45.602  Percent Identity: 22.454

alignment_block:
US-09-135-238B-2 x US-08-378-939-9 ..

Align seg 1/1 to: US-08-378-939-9 from: 1 to: 1617

1 MetAspArgTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLe 17
35 ATGGAC..TGGACCTGGAGGTCTCTCTTTG...GTGGCAGCAGCTAC 78
17 uArgIleLeuProGluValLysVal.....GluGlyG 28
79 AGGTGTCAGTCCAGCATGAGTGGTGTGTCAGTCTGGGGCTGAAGTAAGA 128
28 luLeuGlyGlySerValThrIleLysCysProLeuProGlu..... 41
129 AGCCTGGGTCTCGGTGACGGTCTCTCTGAAGCATCTGGAGGCACCTTC 178
42 MetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGly..... 56
179 AGCACTATGTATCAGCTGGGTGGCAGACAGCCCTCGACAAAGGCTTGA 228
57 .....ThrCysGlyThrValValSerThrThr 66
229 GTGGATGGAGGGATCATCTCTTTTGGTACACCAAGCTACTCACAGA 278
66 snPheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrPro 82
279 ACTTC.....CAGGGCAGAGTCAGGATTACCGCGGACAAA 313
83 ArgLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSe 99
314 TCCACACGACACAGCCACATGGAGCTGACTAGCTGAGATCTGAGGACAC 363
99 rGlyValTyrAlaCysGlyAlaGly.....MetAsnThrAspA 112
364 GCGCGTGATTACTGTGCGACAGATCGCTACAGCGAGGCAGCAATTTGACC 413
112 rgGlyLysThr..GlnLysValThrLeuAsnValHisSerGluTyrGluPr 128
414 GGCGCGGGTGGCTGGTTCGAGCCCTGGGCGCAGGCGACCCCTGGTCACC 463
128 oSerTrpGluGlnProMetProGluThrProLysTrpPheHisLeuP 145
464 GTCTCTCAGCTCCACCAAGGGCCCATCGGTCTTCCCGCTGGCACCTTC 513
145 roTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysPheValThr 161
514 CTCCAAGACACCTCTGGGGGACAGCGGCCCTGGGCTGCCTGG...TCA 560
162 ArgValThrThrProAlaGln.....ArgGlyLysValProProVa 175
561 AGGACTACTTCCCGAAACCGGTGACGGTGTGCTGGAACACTCAGGCGCCCTG 610
175 lHisHisSerSerProThrThrGlnIleThrHisArgProArgValSerA 192
611 ACCAGCGCGTGCACACCTTCCCGCGCTGTCTCTACAGTCTCTCAGGACTCTA 660
```



```
4062 CATGAACGGGTGAAGCAGACGACGATGAGAGAACCTTGAAGTGGATGGAC 4111
70 .....AlaGluTyrLysGly 74
4112 GTATTAACTCCTCAATGGTGGTACTGACTACCAACAGAGTTCAGGAC 4161
75 ArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluVa 91
4162 AAGGCCCTTTAACTGAGCAAGTCAACCAACAGAGCTACATGGAGCT 4211
91 lThrGlnLeuThrGlnSerAspSerGlyValTyrAlaCysGlyAlaGly. 107
4212 CCTCAGCTGACATCTGAGGACTCTGCGACTCTATTACTGTGCAAGAGGCT 4261
108 .....MetAsnThrAspArg...GlyLysThrGlnLysValThrLeu 120
4262 ACTATTACTATCTTGGACTACTGGGTCAAGGAACCTCAGTCACCGTC 4311
121 AsnValHisSerGluTyrGlnProSer...Tyr.GluGluGlnProMetP 136
4312 TCCTCAGCTAGCAACCAAGGCCCATCGGTCTTCCCTGGCACCCTCTC 4361
136 roGluThrProLysTyrPheHisLeuProTyrLeuPheGlnMetProAla 152
4362 CAAGAGCACCTCTGGGGCACAGCGGCCCTGG.....GCTGGC 4399
153 TyrAlaSerSerLysPheValThrArgValThrThrProAlaGln... 168
4400 TGG.....TCAGGACTACTTCCCGAACCGGTG 4428
169 .....ArgGlyLysValProProValHisHisSerSerProThrThrG 183
4429 ACGGTGCTGGAACTCAGCGCCCTGACGCGGTGCGACACTTCCC 4478
183 lNleThrHisArgProArgValSerArgAlaSerSerValAlaGlyAsp 199
4479 GGCTGTCTACAGCTCTCAGGACTCTACTCCTCAGCAGCGTGTGACCG 4528
200 LysProArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLe 216
4529 TGCCT.....CCAGCAGCTGGGCCACCCAGCTACATCTGC 4566
216 uGluGlyLeuLysProGlnThrProSerTyrAsnHisHisThrArgL 233
4567 AAGTGAATCAAGCCCAACCAACAGAGTGGACA.....AGGGAGGGA 4603
233 euHisArgGlnAlaLeuAspTyrGlySerGlnSerGlyArgGluGly 249
4604 .....AGAAAGTGGTGAGAGGCCAGCAC.....AGGGAGGGA 4636
250 GlnGlyPheHisIleLeuIleProThrIleLeuGlyLeuPheLeuLeu 266
4637 .....GGGTGCTGCTGGAGGC 4653
266 aLeuLeuGlyLeuValValLysArgAlaValGluArgArgLysAlaLeu 283
4654 AGGCTCAGCGCTCTGCTGAGCAGC...ATCCCGGCTATGACAGCCCGC 4700
283 erArgArgAlaArgAlaValArgMetArgAlaLeuGluSerSer 299
4701 CCAGGCGCAGAGCAG.....GCCCGCTCTGCC 4729
300 GlnArgProArgGlySerProArgProArgSerGlnAsnAsnIleTyrSe 316
4730 TCTTACCGGAGGCTCTGCCGCCCTCAGTCTCAGGAGAGGCTCT 4779
316 rAlaCysProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGlu 333
4780 TCGGCTTTTCCAGGCTCTGGCAGCAGCAGGCTAGGT.....G 4820
333 laProValProGlyPro.....GlyAlaProLeuProPro 344
4821 CCCCTAACCCAGGCGCTGCACACAAAGGGGCGGTGCTGGGCTCAGACCT 4870
```

```
345 AlaProLeuGlnValSerGluSerProTyrLeuHisAlaProSerLeu 360
4871 GCCAAGACCATATCCGGGAGGACCT.....GCCCTGACCTA 4909
seq_name: /cnc2_6/ptodata/1/ina/5B_COMB.seq:US-08-157-101A-6
seq_documentation_block:
; Sequence 6, Application us/08157101A
; Patent NO. 5808032
; GENERAL INFORMATION:
; APPLICANT: KURIHARA, TATSUYA
; APPLICANT: MATSUKURA, SHIGEKAZU
; APPLICANT: TSURUOKA, NOBUO
; APPLICANT: ARIMA, KENJI
; APPLICANT: NISHIHARA, TATSURO
; TITLE OF INVENTION: ANTI-HBS ANTIBODY GENES AND EXPRESSION
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/157,101A
; FILING DATE: 05-APR-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: TITUS, MARLANA K
; REGISTRATION NUMBER: 35843
; REFERENCE/DOCKET NUMBER: 9437/204199
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3711
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUCH
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-157-101A-6
alignment_scores:
Quality: 125.50 Length: 432
Ratio: 0.668 Gaps: 24
Percent Similarity: 43.519 Percent Identity: 23.843
alignment_block:
US-09-135-238B-2 x US-08-157-101A-6 ..
Align seq 1/1 to: US-08-157-101A-6 from: 1 to: 1576
6 TrpProLeuTyrPheLeuProValSerGlyAlaLeuArgIleLeuProGl 22
17 TGGCTTTTAAAGAGGTGTCAGTGTCTAGGTGCA..... 49
22 uValLysValGluGly.GluLeuGlyGlySerValThrIleLysCysPro 38
50 .....GCTGTGGAGTCTGGGGAGCGGTGTC..... 77
39 LeuProGluMethHisValArgIle..... 46
```



```
78 CAGCCTGGGAGGTCCCTGAGACTCTCTCTGTCAGCCTCTGGATTCAACCTT 127
47 .....TyrLeuCysArgGluMetAlaGlySerGlyThrC 58
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
128 CAGTGAACAATCTATGACTGGGTGCGCCAGCGCTCCAGGCAAGGGGTGG 177
58 ysGlyThrValVal.....SerThrThrAsnPheIleLysAla 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
178 AGTGGGTGGCAGTTATATATATGATGGAAATCATATAATTTACCGCAGAC 227
71 GluTyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPh 87
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
228 TCCGTGAGGGCGGATTCACCATTTCCAGAGACAATTCACAAGAACACACT 277
87 eLeuValGluValThrGlnLeuThrGlnSerAspSerGlyValThrAlaC 104
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
278 GTATCTGGAAGTGAAGAGCTGCAAACTGAGGACACGGGTGTCTATTACT 327
104 ysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysVal..ThrLe 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
328 GT.....ATAGAGATCAAACTTACGGAGTCCACAGA 359
120 uAsnValHisSerGluTyrGluProSerTrpGluGluGlnPrometPro. 136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
360 TTTGACTCTCTGGGCGCCAGGGAACCC...TGGTCACCGTCTCTCTCAGCC 406
137 .....GluThrProLysTrpPhe.....HisLeuPro 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
407 CACCAAGGGGCCATCGGTCTTCCCTGGCACCCCTCTCCAAGAGCACCT 456
146 TyrLeuPheGlnMetProAlaTyrAlaSerSerLysPheValThrAr 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
457 CTGGGGCACAGCGGCCCTGGGTCTGGTCA.....AG 491
162 gValThrThrProAlaGln.....ArgGlyLysValProValh 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
492 GACTACTTCCCGCAACCGGTGACGGTGTCTGTGAACACTCAGGCGCCCTGGC 541
176 isHisSerSerProThrThrGlnIleThrHisArgProArgValSerArg 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
542 CAGCGGGGTGCACACCTTCCCGGTCTCTACAGTCTCTCAGGACTCTACT 591
193 AlaSerValAlaGlyAspLysProArgThrPheLeuProSerThrTh 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
592 CCCTCAGCAGCGTGTGACCGTCCCT.....CCAGCAGCTTG 629
209 rAlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProGlnThrProS 226
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
630 GGCACCCAGACCTACATCTGCAACGTGAATCAACAAGCCAGCAACACAA 679
226 erTyrAsnHisHis.....ThrArgLeu 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
680 GGTGGACAAGAAGTGTGAGCCCAATCTGTGACAAACTCACACATGCC 729
234 HisArgGlnArgAlaLeu.....AspTyrGlySerGlnSerG1 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
730 CACGTGCCAGCACCTGAATCTCTGGGGGACCGTTCAGTCTCTCTCTTC 779
246 yArgGluGly..GlnGlyPheHisIleLeuIleProThrIleLeuGlyLeu 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
780 CCCCCAAAACCAAGGACACCCCTCATGATCTCCGGACCCCTGAGGT... 826
263 PheLeuLeuAlaLeuLeuGly..... 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
827 .....CACATGCGTGGTGGTGGAGCTGAGCCAGCAAGACCCCTGAGG 867
270 .....LeuValValLysArg..... 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
868 TCAAGTTCAACTGGTACGTGGACGCGGTGGAGGTGCATATGCCAAGACA 917
275 .....AlaValGluArgArgLysAlaLeuSerArgArgAla 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
918 AAGCGCGGGAGGAGCAGTACAAGACGACGTACCGGGTGGTCAGCGCTCT 967
```

```
287 ArgArgLeuAla.....ValArgMetArgAl 295
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
968 CACGCTCCTCACAGGACTGGTGTGATGCAAGGAGTACAGTGAAGG 1017
295 aLeuGluSerSerGlnArgProArgGlySerProArgProArgSerGlnA 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1018 TCTCCAAACAAGC.....CCTCCAGCCGCCCATCGAGAAA 1052
312 snAsnIleTyrSer.....AlaCysProArgArgAlaArgGlyAlaAsp 326
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1053 ACCATCTCCAAAGCCAAAGGCGAGCCCGCAGA..... 1084
327 AlaAlaGlyThrGlyGluAlaProValProGly.....Pr 338
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1085 ACCACAGGTGTACCCCTGCCCATCCCGGATGAGCTGACCAAGAAC 1134
338 oglyAlaProLeuProAlaProLeuGlnValSerGlu 351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1135 AGGTCAGCCTGACCTGCTGTGTCAAAGGCTTCTATCCAG 1174
seq_name: /cgn2_5/ptodata/1/lna/6_COMB.seq:US-08-713-118-1
seq_documentation_block:
; Sequence 1, Application US/08713118
; Patent No. 6040436
; GENERAL INFORMATION:
; APPLICANT: Franco, Rodrigo
; APPLICANT: Sun Chen, Ai Ru
; APPLICANT: Suey, David J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING HUMAN NEURONAL
; TITLE OF INVENTION: CALCIUM CHANNEL SUBUNITS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/713,118
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mata, Elizabeth W.
; REGISTRATION NUMBER: 38,236
; REFERENCE/DOCKET NUMBER: ACC96-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7266 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 92..7102
; US-08-713-118-1
```

```
alignment_scores:
Quality: 125.50      Length: 408
Ratio: 0.721        Gaps: 24
Percent Similarity: 42.647      Percent Identity: 23.775
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2000, 10:09:19 : Search time 3690.46 Seconds
(without alignments)
923.671 Million cell updates/sec

Title: US-09-135-238B-1
Perfect score: 1910
Sequence: 1 aaagagtaagcagcgtgc.....ttactctgtccatccttt 1910

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl:*
- 1: gb_bal:*
 - 2: gb_ba2:*
 - 3: gb_om:*
 - 4: gb_ov:*
 - 5: gb_pat:*
 - 6: gb_ph:*
 - 7: gb_pl1:*
 - 8: gb_pl2:*
 - 9: gb_pri:*
 - 10: gb_pr2:*
 - 11: gb_pr3:*
 - 12: gb_ro:*
 - 13: gb_sts:*
 - 14: gb_sy:*
 - 15: gb_un:*
 - 16: em_fun:*
 - 17: em_hum1:*
 - 18: em_hum2:*
 - 19: em_in:*
 - 20: em_om:*
 - 21: em_or:*
 - 22: em_ov:*
 - 23: em_pat:*
 - 24: em_ph:*
 - 25: em_pl:*
 - 26: em_ro:*
 - 27: em_sts:*
 - 28: em_sy:*
 - 29: em_un:*
 - 30: em_vi:*
 - 31: gb_in1:*
 - 32: gb_in2:*
 - 33: gb_in3:*
 - 34: gb_pl3:*
 - 35: gb_pr4:*
 - 36: gb_pr5:*
 - 37: gb_htg1:*
 - 38: gb_htg2:*
 - 39: gb_htg3:*
 - 40: gb_htg4:*
 - 41: gb_htg5:*
 - 42: gb_htg6:*
 - 43: gb_htg7:*

- 44: gb_htg8:*
- 45: gb_htg9:*
- 46: gb_htg10:*
- 47: gb_htg11:*
- 48: gb_htg12:*
- 49: gb_htg13:*
- 50: gb_htg14:*
- 51: gb_htg15:*
- 52: gb_htg16:*
- 53: gb_htg17:*
- 54: em_bai:*
- 55: em_ba2:*
- 56: em_htg1:*
- 57: em_htg2:*
- 58: em_htg3:*
- 59: em_htg4:*
- 60: em_htg5:*
- 61: em_htg6:*
- 62: em_htg7:*
- 63: em_hum3:*
- 64: em_hum4:*
- 65: em_hum5:*
- 66: em_hum6:*
- 67: gb_htg18:*
- 68: gb_htg19:*
- 69: gb_htg20:*
- 70: gb_htg21:*
- 71: gb_htg22:*
- 72: gb_htg23:*
- 73: gb_htg24:*
- 74: gb_htg25:*
- 75: gb_htg26:*
- 76: gb_htg27:*
- 77: gb_htg28:*
- 78: gb_htg29:*
- 79: gb_htg30:*
- 80: gb_htg31:*
- 81: gb_vil:*
- 82: gb_vil2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1300.2	68.1	1339	11	AF057557 Homo sapi
2	745	39.0	221365	47	AC023534 Homo sapi
3	160.6	8.4	58628	71	AC027719 Homo sapi
C 4	106	5.5	221365	47	AC023534 Homo sapi
5	56.2	2.9	7218	5	I66494 Sequence 14
C 6	55.8	2.9	43147	1	AL109663 Streptomy
C 7	54.2	2.8	55482	44	AC022663 Homo sapi
C 8	50.4	2.6	227613	68	AC026852 Homo sapi
C 9	49.6	2.6	224850	77	AC022780 Mus muscu
C 10	47.6	2.5	80507	48	AC025066 Homo sapi
11	47.4	2.5	17578	44	AC023440 Homo sapi
12	47.2	2.5	176705	42	AC018348 Homo sapi
13	47	2.5	2904	9	AB005293 Homo sapi
C 14	46.6	2.4	154566	44	AC008334 Drosophil
C 15	46.4	2.4	159758	68	AC020571 Homo sapi
16	45.4	2.4	168012	53	AC025565 Homo sapi
C 17	45.2	2.4	213063	43	AC009964 Homo sapi
C 18	45.2	2.4	219565	43	AC009974 Homo sapi
19	45	2.4	44169	53	AC025851 Homo sapi
C 20	45	2.4	63827	74	AC037448 Homo sapi
21	45	2.4	76968	44	AC022648 Homo sapi
22	45	2.4	82815	78	AC062001 Homo sapi
23	44.8	2.3	10129	2	AB002081 Deinococc
C 24	44.6	2.3	11440	2	AB002034 Deinococc

25 44.6 2.3 136551 76 AC048354 Homo sapi
26 44.6 2.3 175837 70 AC022045 Homo sapi
c 27 44.4 2.3 145540 9 AP001052 Homo sapi
28 44.4 2.3 179094 46 AC023429 Homo sapi
c 29 44.4 2.3 181490 77 AC009690 Homo sapi
30 44.4 2.3 244466 46 AC009235 Homo sapi
31 44.2 2.3 8113 81 HSBICP4A Li14320 Bovine herp
c 32 44.2 2.3 70542 46 AC022678 Homo sapi
33 44.2 2.3 118831 79 AC064818 Homo sapi
34 44.2 2.3 135301 81 BHV1GEN AC064818 Homo sapi
c 35 44.2 2.3 135301 81 BHV1GEN AC064818 Bovine he
c 36 44 2.3 153367 47 AC021096 AC021096 Bovine he
c 37 43.8 2.3 194828 79 AC006994 AC006994 Homo sapi
38 43.6 2.3 61910 44 AC023002 AC023002 Homo sapi
39 43.6 2.3 78037 74 AC025047 AC025047 Mus muscu
c 40 43.6 2.3 79264 44 AC022707 AC022707 Homo sapi
c 41 43.6 2.3 124328 42 AC016442 AC016442 Homo sapi
c 42 43.6 2.3 236595 48 AC018648 AC018648 Homo sapi
c 43 43.4 2.3 7218 5 I66494 I66494 Sequence 14
c 44 43.4 2.3 65831 76 AC048358 AC048358 Homo sapi
c 45 43.4 2.3 135552 48 AC019251 AC019251 Homo sapi

ALIGNMENTS

RESULT 1
AF057557 1339 bp mRNA PRI 04-JUN-1998
LOCUS Homo sapiens anti-Fas-induced apoptosis (TOSO) mRNA, complete cds.
DEFINITION AF057557.1 GI:3169292
ACCESSION
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1339)
AUTHORS Hitoshi,Y., Lorenz,J., Kitada,S.I., Fisher,J., LaBarge,M.,
Ring,H.Z., Francke,U., Reed,J.C., Kinoshita,S. and Nolan,G.P.
Toso, a cell surface, specific regulator of Fas-induced apoptosis
in T cells
JOURNAL Immunity 8 (4), 461-471 (1998)
MEDLINE 98246048
REFERENCE 2 (bases 1 to 1339)
AUTHORS LaBarge,M. and Hitoshi,Y.
TITLE Direct Submission
JOURNAL Submitted (03-APR-1998) Molecular Pharmacology, Stanford
University, 300 Pasteur Dr., Palo Alto, CA 94304, USA
FEATURES
Location/Qualifiers
1..1339
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q31-q32"
/cell_line="Jurkat"
/cell_type="T cell"
/note="TOSO.2 cDNA; cloned by PCR using primers designed
from the original TOSO.1 cDNA sequence"
1..1339
/gene="TOSO"
20..1192
/gene="TOSO"
/note="transmembrane protein; Ig superfamily member"
/codon_start=1
/product="anti-Fas-induced apoptosis"
/protein_id="AAC18830.1"
/db_xref="GI:3169293"
/translation="MDFWLMPLFLPVSGALRLILPEVKVEGLGSGVTIKPLPEMHV
RIYLCHREMGSGCTGVSTNFIKAEYKGRVTLKQYPRKNLFLVEVTLQTFSDSGVY
ACAGNMTRDKTKVTLNPFSEYEPFSEWQPMPEPKWFLPYLPFOMPAVSSKFP
TRYVTPAQRGVPPVHHSSPTQILTRPRVSRSSVAGDKPRFLPSTASKISALEG
LLAPQTPSYNHHTRLRQALDVGSGSGREGQGHILFIPLGLFLALLGLVKKRV

gene

CDS

ERRKALSRRRRLAVRMRALESSQRPRGRPRSRQNNIYSACPRRAGADAAGTGEAP
VPGGAPLPAPQVSESPWLHAPSLKTSCEVSLYHQPAAMMEDSDSDDIYNPA"
127
variation /gene="TOSO"
/note="occurs in TOSO.1 cDNA"
/replace="a"
1015
variation /gene="TOSO"
/note="occurs in TOSO.1 cDNA"
/replace="g"
BASE COUNT 309 a 432 c 340 g 258 t
ORIGIN
Query Match 68.1%; Score 1300.2; DB 11; Length 1339;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1335; Conservative 0; Mismatches 3; Indels 3; Gaps 3;
QY 55 ttgcactctagaaggacaaatggactctggccttggccactttactctgcagatc 114
Db 1 TTGCACTCTAGAGGGACAATGGACTTCTGGCTTTGGCCACTTTACTTCTGCAGTATC 60
QY 115 agggccctgaggatcctcccagaagtaaaagtagagggggagctggcgagatcagttac 174
Db 61 GGGGCCCTGAGGATCCTCCAGAAAGTAAGGTAGAGGGGAGCTGGGGGATCAGTTAC 120
QY 175 catcaaatgcccacttctgaaatgcagtgagagatatctgtgccggagatgctg 234
Db 121 CATCAAGTGCCCACTTCTCGAAATGCATGTGAGGATATATCTGTGCCGGAGATGCTGG 180
QY 235 atctggaacatgtgtaccgtggtatccaccaccaacttcacaggcagaatacaagg 294
Db 181 ATCTGGAACATGTGGTACCGTGGTATCCACCACCACTTCAICAAAGCAGAATACAAAGG 240
QY 295 ccgagttactgaagcaatacccaacgaagaatctgttctcctagtgaggttaacacag 354
Db 241 CGAGTTTACTGTGAACCAATACCCACGCAAGAATCTGTTCCTAGTGGAGGTAAACAGCT 300
QY 355 gacagaagtgacagcgaggtctatgctcgagcgagcgagcatgaacacagaccgggaaa 414
Db 301 GACAGAAAGTGACAGCGGAGTCTATGCTCGGAGCGGCGATGAAACACACACCGGGGAAA 360
QY 415 gaccagaagtcacccctgaaatgtccacagtgaaatcacagagccatctgggaagagcagcc 474
Db 361 GACCAGAAAGTCACCCCTGAATGTCCAGTGAATACGAGGCCATCATGGGAAGAGCAGCC 420
QY 475 aatgctgagactccaaaatgttccatctgccctattgttccagatgctgcataatgc 534
Db 421 AATGCTGAGACTCCAAAATGGTTTCATCTGCCCTATTGTTTCCAGATGCTGCATATGC 480
QY 535 cagttcttccaaattcgttaaccagagttaccacaccagctcaaaaggggcaaggtccctcc 594
Db 481 CAGTTCTTCCAAAATTCGTAACCAAGATTACCAACAGCTCAAAAGGGGCAAGTCCCTCC 540
QY 595 agttcacactctctcccccaccccaaatcccccaccccgctcaggttcagagcagc 654
Db 541 AGTTTCAACACTCTCTCCCCACCAACCCCAAAATCAACCCACCGCTCGAGTGTCCAGAGCATC 600
QY 655 ttccagtagcaggtgacaagccccgaactctctgcatccactacagctcctcaaaatctc 714
Db 601 TTCACTAGCAGGTGACAAAGCCGCAAGCTTCTTCCATCCACTACAGCTCAAAAATCTC 660
QY 715 agctctgagggctgctcaagcccgagcgccagctacacacacacacagaggtgca 774
Db 661 AGCTCTGGAGGGGCTCTCAGCCCCAGAGCGCCAGCTACACACCAACACAGGCTGCA 720
QY 775 caggcagagagcactggactatggtcacagctctggagggaaggaaggaaggtttccat 834
Db 721 CAGGCAGAGAGCAGTGGACTATGCTCAGTCTGGAGGGAAGGCCCAAGGATTTACAT 780
QY 835 cctaatccccaccactctggcctttctctgctggcaactctctggccttctggcctggtgaaag 894
Db 781 CCTGATCCCGACCATCTCTGGGCGCTTTTCTCTGCTGGCCTTCTTGGGGCTGTGTGAAAG 840

* 34913 36817: contig of 1905 bp in length
* 36818 36917: gap of 100 bp
* 36918 38575: contig of 1658 bp in length
* 38576 38675: gap of 100 bp
* 38676 41309: contig of 2634 bp in length
* 41310 41409: gap of 100 bp
* 41410 43344: contig of 1935 bp in length
* 43345 43444: gap of 100 bp
* 43445 45853: contig of 2409 bp in length
* 45854 45953: gap of 100 bp
* 45954 48434: contig of 2481 bp in length
* 48435 48534: gap of 100 bp
* 48535 51026: contig of 2492 bp in length
* 51027 51126: gap of 100 bp
* 51127 54369: contig of 3243 bp in length
* 54370 54469: gap of 100 bp
* 54470 56661: contig of 2192 bp in length
* 56662 56761: gap of 100 bp
* 56762 59097: contig of 2336 bp in length
* 59098 59197: gap of 100 bp
* 59198 61551: contig of 2354 bp in length
* 61552 61651: gap of 100 bp
* 61652 64257: contig of 2606 bp in length
* 64258 64357: gap of 100 bp
* 64358 67511: contig of 3154 bp in length
* 67512 67611: gap of 100 bp
* 67612 70329: contig of 2718 bp in length
* 70330 70429: gap of 100 bp
* 70430 73572: contig of 3143 bp in length
* 73573 73672: gap of 100 bp
* 73673 77203: contig of 3531 bp in length
* 77204 77303: gap of 100 bp
* 77304 80599: contig of 3296 bp in length
* 80600 80699: gap of 100 bp
* 80700 84046: contig of 3347 bp in length
* 84047 84146: gap of 100 bp
* 84147 88357: contig of 4211 bp in length
* 88358 88457: gap of 100 bp
* 88458 92454: contig of 3997 bp in length
* 92455 92554: gap of 100 bp
* 92555 96892: contig of 4338 bp in length
* 96893 96992: gap of 100 bp
* 96993 101334: contig of 4342 bp in length
* 101335 101434: gap of 100 bp
* 101435 105438: contig of 4004 bp in length
* 105439 105538: gap of 100 bp
* 105539 110413: contig of 4875 bp in length
* 110414 110513: gap of 100 bp
* 110514 115994: contig of 5481 bp in length
* 115995 116094: gap of 100 bp
* 116095 120447: contig of 4353 bp in length
* 120448 120547: gap of 100 bp
* 120548 124483: contig of 3936 bp in length
* 124484 124583: gap of 100 bp
* 124584 130249: contig of 5666 bp in length
* 130250 130349: gap of 100 bp
* 130350 135462: contig of 5113 bp in length
* 135463 135562: gap of 100 bp
* 135563 139847: contig of 4285 bp in length
* 139848 139947: gap of 100 bp
* 139948 146731: contig of 6784 bp in length
* 146732 146831: gap of 100 bp
* 146832 151469: contig of 4638 bp in length
* 151470 151569: gap of 100 bp
* 151570 159901: contig of 8332 bp in length
* 159902 160001: gap of 100 bp
* 160002 168308: contig of 8307 bp in length
* 168309 168408: gap of 100 bp
* 168409 176195: contig of 7787 bp in length
* 176196 176295: gap of 100 bp
* 176296 189303: contig of 13008 bp in length
* 189304 189403: gap of 100 bp
* 189404 202135: contig of 12732 bp in length

* 202136 202235: gap of 100 bp
* 202236 221365: contig of 19130 bp in length.
FEATURES
source
1..221365
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPC1-11 Human Male BAC"
1..1108
/note="assembly_fragment"
1209..2292
/note="assembly_fragment"
2393..3673
/note="assembly_fragment"
3774..4947
/note="assembly_fragment"
5048..6104
/note="assembly_fragment"
6205..7259
/note="assembly_fragment"
7360..8667
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
Query Match 39.0%; Score 745; DB 47; Length 221365;
Best Local Similarity 98.2%; Pred. No. 7.1e-178;
Matches 785; Conservative 0; Mismatches 10; Indels 4; Gaps 3;
QY 1114 caagatctgaatctccctgctccatccccatctctgaagaccagctgaatcgtg 1173
DB 58030 CAGGTGTCGTAATCTCCCTGGCTCCATGCCCATCTCTGAAGACCAGCTGTAATACGG 58089
QY 1174 agccttaccaccagctgcgcgcctgatgagacagtgattcagatgactacatcaat 1233
DB 58090 AGCCTTACCACCAGCCTGCCCATGATGAGGACAGTAGTTCAGATGACTACATCAAT 58149
QY 1234 gtctctgcctgacaactccccagctatcccccaacccagctcgagctgtggtgcgaag 1293
DB 58150 GTTCTGCTGACAACTCCCCAGCTATCCCCCAACCCAGGCTCGGACTGTGTGCCAAG 58209
QY 1294 gagtctcatctatctgctgatgtccaaatcctgctcattgttctctgagagccctcatca 1353
DB 58210 GAGTCTCATCTATCTGCTGATGTCCAACTGCTTCATGTGTCTCAGAGCCCTCATCA 58269
QY 1354 cttccatgccccatctctgactccccatctctgtgacctgagctgagctgtctgt 1413
DB 58270 CTTCCCATGCCCATCTCGACTCCCATCCCATATCTGT-GCCCTGAGCATGGCTCTG 58328
QY 1414 cccccaggtcgtcttgacaccccttgacagccccctgtagttgacaggtgaagctgtagcca 1473
DB 58329 CCCCCAGGTCTGTTGCACACCTTGGCAGCCCCCTGTAGTTGACAGGTAAGCTGTAGGCA 58388
QY 1474 tgttagacaattgtcccaatgccattgctcttcccaagcgtcgaaacagactgtggg 1533
DB 58389 TGTAGAGCAATGTGCCAATGCCATGTCTTCCATTCCTCCAAAGCCGTCGAACAGCTGTGG 58448
QY 1534 atttcagagtggtttcttccatgcttttgacacaggggtgtgtgtgctgcagagctctag 1593
DB 58449 ATTTGCAGAGTGTCTTCTTCCATGTCTTTGACCACAGGGT-TGTTGCTGCCCAGGCTCTAG 58507
QY 1594 atcaatggcatcagcgtggggcagagcagatgatttctctcgggcatcttccacagg 1653
DB 58508 ATCATATGGCATCAGGCTGGGGCAGAGCATAGCTATTGTCTCGGGCATCTTCCACAGG 58567
QY 1654 ttgggtttacacaaatagaagctcttctgctgagttatgtgaagctctcagagcccat 1713
DB 58568 TTGGGTCTTACACAAATAGAAGGCTCTTGTCTCTGAGTTATGTGTAATGCCATGCCATCCAT 58627
QY 1714 ggactaagcaggggtctgggtat--aaacactcctggaaacgcttttgcctctgatccaaat 1771
DB 58628 GGACTAAGCAGGGGCTGGTATATAAAACACTCTCTGGAACGCCCTTGGCCTGTGATCAAAAT 58687
QY 1772 gttagcaactgtctaagaacgctctacttatctcaagttctatgcttaagaaagcaatttct 1831
DB 58688 GTTAGCACTGTCTAGTGAAGCTCTACTTATCTCAAGTTCTATGCTTAAGGCAATTTATCT 58747


```

OY 1832 tgatgtgataaaacaaactattagcaagatgcatatataatcattctctt 1891
LOCUS |||||||
Db 58748 TGATGTGATGATAAACCAACTATTATTAGCAAGATGATGATATATATCCATAAAATCTCTT 58807
OY 1892 tactctgtctccatctctt 1910
LOCUS |||||||
Db 58808 TACTCTGTCTCCATCACTT 58826

RESULT 3
AC027719/c
DEFINITION Homo sapiens chromosome 8 clone RP11-258P9 map 8, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC027719
VERSION AC027719.1 GI:7382531
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 58628)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 8, clone RP11-258P9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 58628)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F.,
Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galegan, J., Gardyna, S., Glende, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Weldrim, J., Maneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, B., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirtrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7449
Center clone name: 258_P_9
-----
* NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will

```

```

* be preserved.
* 1 671: contig of 671 bp in length
* 672 771: gap of 100 bp
* 772 1469: contig of 698 bp in length
* 1470 1569: gap of 100 bp
* 1570 2261: contig of 692 bp in length
* 2262 2361: gap of 100 bp
* 2362 3051: contig of 690 bp in length
* 3052 3151: gap of 100 bp
* 3152 3816: contig of 665 bp in length
* 3817 3916: gap of 100 bp
* 3917 4597: contig of 681 bp in length
* 4598 4697: gap of 100 bp
* 4698 5395: contig of 698 bp in length
* 5396 5495: gap of 100 bp
* 5496 6185: contig of 690 bp in length
* 6186 6285: gap of 100 bp
* 6286 6969: contig of 684 bp in length
* 6970 7069: gap of 100 bp
* 7070 7780: contig of 711 bp in length
* 7781 7880: gap of 100 bp
* 7881 8590: contig of 710 bp in length
* 8591 8690: gap of 100 bp
* 8691 9372: contig of 682 bp in length
* 9373 9472: gap of 100 bp
* 9473 10162: contig of 690 bp in length
* 10163 10262: gap of 100 bp
* 10263 10944: contig of 682 bp in length
* 10945 11044: gap of 100 bp
* 11045 11747: contig of 703 bp in length
* 11748 11847: gap of 100 bp
* 11848 12547: contig of 700 bp in length
* 12548 12647: gap of 100 bp
* 12648 13336: contig of 689 bp in length
* 13337 13436: gap of 100 bp
* 13437 14103: contig of 667 bp in length
* 14104 14203: gap of 100 bp
* 14204 14872: contig of 669 bp in length
* 14873 14972: gap of 100 bp
* 14973 15682: contig of 710 bp in length
* 15683 15782: gap of 100 bp
* 15783 16491: contig of 709 bp in length
* 16492 16591: gap of 100 bp
* 16592 17290: contig of 699 bp in length
* 17291 17390: gap of 100 bp
* 17391 18088: contig of 698 bp in length
* 18089 18188: gap of 100 bp
* 18189 18883: contig of 695 bp in length
* 18884 18983: gap of 100 bp
* 18984 19671: contig of 688 bp in length
* 19672 19771: gap of 100 bp
* 19772 20413: contig of 642 bp in length
* 20414 20513: gap of 100 bp
* 20514 21196: contig of 683 bp in length
* 21197 21296: gap of 100 bp
* 21297 21992: contig of 696 bp in length
* 21993 22092: gap of 100 bp
* 22093 22804: contig of 712 bp in length
* 22805 22904: gap of 100 bp
* 22905 23604: contig of 700 bp in length
* 23605 23704: gap of 100 bp
* 23705 24402: contig of 698 bp in length
* 24403 24502: gap of 100 bp
* 24503 25191: contig of 689 bp in length
* 25192 25291: gap of 100 bp
* 25292 25974: contig of 683 bp in length
* 25975 26074: gap of 100 bp
* 26075 26769: contig of 695 bp in length
* 26770 26869: gap of 100 bp
* 26870 27572: contig of 703 bp in length
* 27573 27672: gap of 100 bp
* 27673 28373: contig of 701 bp in length
* 28374 28473: gap of 100 bp

```

```

* 28474 29165: contig of 692 bp in length
* 29166 29265: gap of 100 bp
* 29266 29956: contig of 691 bp in length
* 29957 30056: gap of 100 bp
* 30057 30741: contig of 685 bp in length
* 30742 30841: gap of 100 bp
* 30842 31519: contig of 678 bp in length
* 31520 31619: gap of 100 bp
* 31620 32323: contig of 704 bp in length
* 32324 32423: gap of 100 bp
* 32424 33110: contig of 687 bp in length
* 33111 33210: gap of 100 bp
* 33211 33902: contig of 692 bp in length
* 33903 34002: gap of 100 bp
* 34003 34708: contig of 706 bp in length
* 34709 34808: gap of 100 bp
* 34809 35500: contig of 692 bp in length
* 35501 35600: gap of 100 bp
* 35601 36300: contig of 700 bp in length
* 36301 36400: gap of 100 bp
* 36401 37097: contig of 697 bp in length
* 37098 37197: gap of 100 bp
* 37198 37900: contig of 703 bp in length
* 37901 38000: gap of 100 bp
* 38001 38695: contig of 695 bp in length
* 38696 38795: gap of 100 bp
* 38796 39487: contig of 692 bp in length
* 39488 39587: gap of 100 bp
* 39588 40291: contig of 704 bp in length
* 40292 40391: gap of 100 bp
* 40392 41101: contig of 710 bp in length
* 41102 41886: contig of 685 bp in length
* 41887 41986: gap of 100 bp
* 41987 42686: contig of 700 bp in length
* 42687 42786: gap of 100 bp
* 42787 43484: contig of 698 bp in length
* 43485 44276: contig of 692 bp in length
* 44277 44376: gap of 100 bp
* 44377 45078: contig of 702 bp in length
* 45079 45178: gap of 100 bp
* 45179 45869: contig of 691 bp in length
* 45870 45969: gap of 100 bp
* 45970 46664: contig of 695 bp in length
* 46665 46764: gap of 100 bp
* 46765 47456: contig of 692 bp in length
* 47457 47556: gap of 100 bp
* 47557 48249: contig of 693 bp in length
* 48250 48349: gap of 100 bp
* 48350 49042: contig of 693 bp in length
* 49043 49142: gap of 100 bp
* 49143 49843: contig of 701 bp in length
* 49844 49943: gap of 100 bp
* 49944 50640: contig of 697 bp in length
* 50641 50740: gap of 100 bp
* 50741 51438: contig of 698 bp in length
* 51439 51538: gap of 100 bp
* 51539 52231: contig of 693 bp in length
* 52232 52331: gap of 100 bp
* 52332 53020: contig of 689 bp in length
* 53021 53120: gap of 100 bp
* 53121 53857: contig of 737 bp in length
* 53858 53957: gap of 100 bp
* 53958 54652: contig of 695 bp in length
* 54653 54752: gap of 100 bp

Query Match      8.4%; Score 160.6; DB 71; Length 58628;
Best Local Similarity 97.6%; Pred. No. 6e-30;
Matches 163; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 102 tctgtccatcagggccctgaggtatctccagaaagtaagtagagggagctgg 161
|||||

```

```

Db 40603 TCCCTGCAGTATCAGGGCCCTCAGGATCCTCCAGAGTAAGTAGAGGGAGGCTGG 40544
QY 162 gggatcagttaccatcaaatgccacttcctgaaatgcattgaggtatatctgtgcc 221
|||||
Db 40543 CGCGATCAGTTACCATCAAGTCCCACTTCCTCGAAATGCATGTGAGATATATCTGTGCC 40484
QY 222 gggagatggctgagatctggaacatggtaccggtggtatccaccacc 268
|||||
Db 40483 GGGAGATGGCTGGATCTGGAACATGTGGTACCGTGTATCCACCACC 40437

RESULT 4
AC023534/c
LOCUS
DEFINITION
AC023534 Homo sapiens clone RP11-462N18, *** SEQUENCING IN PROGRESS ***, 61
unordered pieces.
AC023534
VERSION
AC023534.2 GI:7143452
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 221365)
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE
Homo sapiens chromosome, clone RP11-462N18
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 221365)
AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Bouckgaert,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeAerlando,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Gadigan,P., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,K., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Larcocque,R., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisan,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Willson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2000 this sequence version replaced gi:6978210.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center Project name: L5244
Center Clone name: 462_N_18
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

* 1 1108: contig of 1108 bp in length
* 1109 1208: gap of 100 bp
* 1209 2292: contig of 1084 bp in length
* 2293 2392: gap of 100 bp
* 2393 3673: contig of 1281 bp in length
* 3674 3773: gap of 100 bp
* 3774 4947: contig of 1174 bp in length
* 4948 5047: gap of 100 bp
* 5048 6104: contig of 1057 bp in length
* 6105 6204: gap of 100 bp
* 6205 7259: contig of 1055 bp in length
* 7260 7359: gap of 100 bp
* 7360 8667: contig of 1308 bp in length
* 8668 8767: gap of 100 bp
* 8768 10566: contig of 1799 bp in length
* 10567 10666: gap of 100 bp
* 10667 12173: contig of 1507 bp in length
* 12174 12273: gap of 100 bp
* 12274 13643: contig of 1370 bp in length
* 13644 13743: gap of 100 bp
* 13744 15215: contig of 1472 bp in length
* 15216 15315: gap of 100 bp
* 15316 16609: contig of 1294 bp in length
* 16610 16709: gap of 100 bp
* 16710 18085: contig of 1376 bp in length
* 18086 18185: gap of 100 bp
* 18186 19668: contig of 1483 bp in length
* 19669 19768: gap of 100 bp
* 19769 22083: contig of 2315 bp in length
* 22084 22183: gap of 100 bp
* 22184 24052: contig of 1869 bp in length
* 24053 24152: gap of 100 bp
* 24153 25902: contig of 1750 bp in length
* 25903 26002: gap of 100 bp
* 26003 27639: contig of 1637 bp in length
* 27640 27739: gap of 100 bp
* 27740 29377: contig of 1638 bp in length
* 29378 29477: gap of 100 bp
* 29478 30910: contig of 1433 bp in length
* 30911 31010: gap of 100 bp
* 31011 31225: contig of 215 bp in length
* 31226 31325: gap of 100 bp
* 31326 32890: contig of 1565 bp in length
* 32891 32990: gap of 100 bp
* 32991 34812: contig of 1822 bp in length
* 34813 34912: gap of 100 bp
* 34913 36817: contig of 1905 bp in length
* 36818 36917: gap of 100 bp
* 36918 38575: contig of 1658 bp in length
* 38576 38675: gap of 100 bp
* 38676 41309: contig of 2634 bp in length
* 41310 41409: gap of 100 bp
* 41410 43344: contig of 1935 bp in length
* 43345 43444: gap of 100 bp
* 43445 45853: contig of 2409 bp in length
* 45854 45953: gap of 100 bp
* 45954 48434: contig of 2481 bp in length
* 48435 48534: gap of 100 bp
* 48535 51026: contig of 2492 bp in length
* 51027 51126: gap of 100 bp
* 51127 54369: contig of 3243 bp in length
* 54370 54469: gap of 100 bp
* 54470 56661: contig of 2192 bp in length
* 56662 56761: gap of 100 bp
* 56762 59097: contig of 2336 bp in length
* 59098 59197: gap of 100 bp
* 59198 61551: contig of 2354 bp in length
* 61552 61651: gap of 100 bp
* 61652 64257: contig of 2606 bp in length
* 64258 64357: gap of 100 bp
* 64358 67511: contig of 3154 bp in length
* 67512 67611: gap of 100 bp
* 67612 70329: contig of 2718 bp in length

* 70330 70429: gap of 100 bp
* 70430 73572: contig of 3143 bp in length
* 73573 73672: gap of 100 bp
* 73673 77203: contig of 3531 bp in length
* 77204 77303: gap of 100 bp
* 77304 80599: contig of 3296 bp in length
* 80600 80699: gap of 100 bp
* 80700 84046: contig of 3347 bp in length
* 84047 84146: gap of 100 bp
* 84147 88357: contig of 4211 bp in length
* 88358 88457: gap of 100 bp
* 88458 92454: contig of 3997 bp in length
* 92455 92554: gap of 100 bp
* 92555 96892: contig of 4338 bp in length
* 96893 96992: gap of 100 bp
* 96993 101334: contig of 4342 bp in length
* 101335 101434: gap of 100 bp
* 101435 105438: contig of 4004 bp in length
* 105439 105538: gap of 100 bp
* 105539 110413: contig of 4875 bp in length
* 110414 110513: gap of 100 bp
* 110514 115994: contig of 5481 bp in length
* 115995 116094: gap of 100 bp
* 116095 120447: contig of 4353 bp in length
* 120448 120547: gap of 100 bp
* 120548 124483: contig of 3936 bp in length
* 124484 124583: gap of 100 bp
* 124584 130249: contig of 5666 bp in length
* 130250 130349: gap of 100 bp
* 130350 135462: contig of 5113 bp in length
* 135463 135562: gap of 100 bp
* 135563 139847: contig of 4285 bp in length
* 139848 139947: gap of 100 bp
* 139948 146731: contig of 6784 bp in length
* 146732 146831: gap of 100 bp
* 146832 151469: contig of 4638 bp in length
* 151470 151569: gap of 100 bp
* 151570 159901: contig of 8332 bp in length
* 159902 160001: gap of 100 bp
* 160002 168308: contig of 8307 bp in length
* 168309 168408: gap of 100 bp
* 168409 176195: contig of 7787 bp in length
* 176196 176295: gap of 100 bp
* 176296 189303: contig of 13008 bp in length
* 189304 189403: gap of 100 bp
* 189404 202135: contig of 12732 bp in length
* 202136 202235: gap of 100 bp
* 202236 221365: contig of 19130 bp in length.

FEATURES

source

1. 221365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-462N18"

/clone_lib="RPC1-11 Human Male BAC"
1. 11108
/note="assembly_fragment"

misc_feature

1209..2292
/note="assembly_fragment"

misc_feature

2393..3673
/note="assembly_fragment"

misc_feature

3774..4947
/note="assembly_fragment"

misc_feature

5048..6104
/note="assembly_fragment"

misc_feature

6205..7259
/note="assembly_fragment"

misc_feature

7360..8667
/note="assembly_fragment"

Query Match

Best Local Similarity 5.5%; Score 106; DB 47; Length 221365;

Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 gagtaagcagcgtgtctccatccctctctagggggtctgttgatggacctgcactcta 64


```

LOCUS       AC022663       55482 bp      DNA           HTG           06-FEB-2000
DEFINITION  Homo sapiens clone RP11-2905, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC022663
VERSION     AC022663.1 GI:6910643
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 55482)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-2905
JOURNAL
REFERENCE   2 (bases 1 to 55482)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
             Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
             Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
             Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
             DeArelano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
             Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
             Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
             Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
             Landers,T., Lehoczyk,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
             Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
             McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
             Norman,C.H., O'Connor,T., O'Donnell,P., Oliviar,T.M., Peterson,K.,
             Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
             Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
             Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
             Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
             Zimmer,A. and Zody,M.
             Direct Submission
TITLE       Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
JOURNAL     Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             ----- Genome Center
             Center: Whitehead Institute/ MIT Center for Genome Research
             Center code: WIBR
             Web site: http://www-seq.wi.mit.edu
             Contact: sequence_submissions@genome.wi.mit.edu
             ----- Project Information
             Center project name: L4854
             Center clone name: 29_Q_5
             -----
             * NOTE: This record contains 77 individual
             * sequencing reads that have not been assembled into
             * contigs. Runs of N are used to separate the reads
             * and the order in which they appear is completely
             * arbitrary. Low-pass sequence sampling is useful for
             * identifying clones that may be gene-rich and allows
             * overlap relationships among clones to be deduced.
             * However, it should not be assumed that this clone
             * will be sequenced to completion. In the event that
             * the record is updated, the accession number will
             * be preserved.
             *
             * 1 734: contig of 734 bp in length
             * gap of unknown length
             * 735 1456: contig of 722 bp in length
             *
             * 1457 2152: gap of unknown length
             *
             * 2153 2848: contig of 696 bp in length
             *
             * 2849 3545: contig of 697 bp in length
             *
             * 3546 4228: gap of unknown length
             *
             * 4229 4948: contig of 720 bp in length
             *
             * 4949 5662: contig of 714 bp in length
             *
             * gap of unknown length
             *
             * 5663 6371: contig of 709 bp in length
             * gap of unknown length
             * 6372 7097: contig of 726 bp in length
             * gap of unknown length
             * 7098 7835: contig of 738 bp in length
             * gap of unknown length
             * 7836 8564: contig of 729 bp in length
             * gap of unknown length
             * 8565 9315: contig of 751 bp in length
             * gap of unknown length
             * 9316 10129: contig of 814 bp in length
             * gap of unknown length
             * 10130 10894: contig of 765 bp in length
             * gap of unknown length
             * 10895 11676: contig of 782 bp in length
             * gap of unknown length
             * 11677 12391: contig of 715 bp in length
             * gap of unknown length
             * 12392 13097: contig of 706 bp in length
             * gap of unknown length
             * 13098 13807: contig of 710 bp in length
             * gap of unknown length
             * 13808 14532: contig of 725 bp in length
             * gap of unknown length
             * 14533 15268: contig of 736 bp in length
             * gap of unknown length
             * 15269 16005: contig of 737 bp in length
             * gap of unknown length
             * 16006 16688: contig of 683 bp in length
             * gap of unknown length
             * 16689 17402: contig of 714 bp in length
             * gap of unknown length
             * 17403 18116: contig of 714 bp in length
             * gap of unknown length
             * 18117 18830: contig of 714 bp in length
             * gap of unknown length
             * 18831 19543: contig of 713 bp in length
             * gap of unknown length
             * 19544 20250: contig of 707 bp in length
             * gap of unknown length
             * 20251 20985: contig of 735 bp in length
             * gap of unknown length
             * 20986 21706: contig of 721 bp in length
             * gap of unknown length
             * 21707 22447: contig of 741 bp in length
             * gap of unknown length
             * 22448 23176: contig of 729 bp in length
             * gap of unknown length
             * 23177 23907: contig of 731 bp in length
             * gap of unknown length
             * 23908 24626: contig of 719 bp in length
             * gap of unknown length
             * 24627 25337: contig of 711 bp in length
             * gap of unknown length
             * 25338 26059: contig of 722 bp in length
             * gap of unknown length
             * 26060 26774: contig of 715 bp in length
             * gap of unknown length
             * 26775 27484: contig of 710 bp in length
             * gap of unknown length
             * 27485 28211: contig of 727 bp in length
             * gap of unknown length
             * 28212 28936: contig of 725 bp in length
             * gap of unknown length
             * 28937 29621: contig of 685 bp in length
             * gap of unknown length
             * 29622 30344: contig of 723 bp in length
             * gap of unknown length
             * 30345 31053: contig of 709 bp in length
             * gap of unknown length
             * 31054 31769: contig of 716 bp in length
             * gap of unknown length
             * 31770 32478: contig of 709 bp in length
             * gap of unknown length

```


* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 762: contig of 762 bp in length
* 763 862: gap of 100 bp
* 863 1656: contig of 794 bp in length
* 1657 1756: gap of 100 bp
* 1757 2583: contig of 807 bp in length
* 2564 2663: gap of 100 bp
* 2664 3449: contig of 786 bp in length
* 3450 3549: gap of 100 bp
* 3550 4325: contig of 776 bp in length
* 4326 4425: gap of 100 bp
* 4426 5227: contig of 802 bp in length
* 5228 5327: gap of 100 bp
* 5328 6139: contig of 812 bp in length
* 6140 6239: gap of 100 bp
* 6240 7044: contig of 805 bp in length
* 7045 7144: gap of 100 bp
* 7145 7949: contig of 805 bp in length
* 7950 8049: gap of 100 bp
* 8050 8868: contig of 819 bp in length
* 8869 8968: gap of 100 bp
* 8969 9801: contig of 833 bp in length
* 9802 9901: gap of 100 bp
* 9902 10610: contig of 703 bp in length
* 10611 10710: gap of 100 bp
* 10711 11507: contig of 797 bp in length
* 11508 11607: gap of 100 bp
* 11608 12409: contig of 802 bp in length
* 12410 12509: gap of 100 bp
* 12510 13307: contig of 798 bp in length
* 13308 13407: gap of 100 bp
* 13408 14187: contig of 780 bp in length
* 14188 14287: gap of 100 bp
* 14288 15062: contig of 775 bp in length
* 15063 15162: gap of 100 bp
* 15163 15978: contig of 816 bp in length
* 15979 16078: gap of 100 bp
* 16079 16875: contig of 797 bp in length
* 16876 16975: gap of 100 bp
* 16976 17676: contig of 701 bp in length
* 17677 17776: gap of 100 bp
* 17777 18581: contig of 805 bp in length
* 18582 18681: gap of 100 bp
* 18682 19442: contig of 761 bp in length
* 19443 19542: gap of 100 bp
* 19543 20326: contig of 784 bp in length
* 20327 20426: gap of 100 bp
* 20427 21241: contig of 815 bp in length
* 21242 21341: gap of 100 bp
* 21342 22132: contig of 791 bp in length
* 22133 22232: gap of 100 bp
* 22233 23027: contig of 795 bp in length
* 23028 23127: gap of 100 bp
* 23128 23916: contig of 789 bp in length
* 23917 24016: gap of 100 bp
* 24017 24826: contig of 810 bp in length
* 24827 24926: gap of 100 bp
* 24927 25738: contig of 812 bp in length
* 25739 25838: gap of 100 bp
* 25839 26646: contig of 808 bp in length
* 26647 26746: gap of 100 bp
* 26747 27556: contig of 810 bp in length
* 27557 27656: gap of 100 bp
* 27657 28460: contig of 804 bp in length
* 28461 28560: gap of 100 bp
* 28561 29364: contig of 804 bp in length
* 29365 29464: gap of 100 bp
* 29465 30259: contig of 795 bp in length
* 30260 30359: gap of 100 bp

* 30360 31149: contig of 790 bp in length
* 31150 31249: gap of 100 bp
* 31250 32058: contig of 809 bp in length
* 32059 32158: gap of 100 bp
* 32159 32968: contig of 810 bp in length
* 32969 33068: gap of 100 bp
* 33069 33870: contig of 802 bp in length
* 33871 33970: gap of 100 bp
* 33971 34769: contig of 799 bp in length
* 34770 34869: gap of 100 bp
* 34870 35682: contig of 813 bp in length
* 35683 35782: gap of 100 bp
* 35783 36585: contig of 803 bp in length
* 36586 36685: gap of 100 bp
* 36686 37501: contig of 816 bp in length
* 37502 37601: gap of 100 bp
* 37602 38393: contig of 792 bp in length
* 38394 38493: gap of 100 bp
* 38494 39303: contig of 810 bp in length
* 39304 39403: gap of 100 bp
* 39404 40198: contig of 795 bp in length
* 40199 40298: gap of 100 bp
* 40299 41092: contig of 794 bp in length
* 41093 41192: gap of 100 bp
* 41193 41982: contig of 790 bp in length
* 41983 42082: gap of 100 bp
* 42083 42908: contig of 826 bp in length
* 42909 43008: gap of 100 bp
* 43009 43788: contig of 780 bp in length
* 43789 43888: gap of 100 bp
* 43889 44662: contig of 774 bp in length
* 44663 44762: gap of 100 bp
* 44763 45556: contig of 794 bp in length
* 45557 45656: gap of 100 bp
* 45657 46462: contig of 806 bp in length
* 46463 46562: gap of 100 bp
* 46563 47367: contig of 805 bp in length
* 47368 47467: gap of 100 bp
* 47468 48267: contig of 800 bp in length
* 48268 48367: gap of 100 bp
* 48368 49156: contig of 789 bp in length
* 49157 49256: gap of 100 bp
* 49257 50060: contig of 804 bp in length
* 50061 50160: gap of 100 bp
* 50161 50952: contig of 792 bp in length
* 50953 51052: gap of 100 bp
* 51053 51860: contig of 808 bp in length
* 51861 51960: gap of 100 bp
* 51961 52753: contig of 793 bp in length
* 52754 52853: gap of 100 bp
* 52854 53650: contig of 797 bp in length
* 53651 53750: gap of 100 bp
* 53751 54532: contig of 782 bp in length
* 54533 54632: gap of 100 bp
* 54633 55430: contig of 798 bp in length
* 55431 55530: gap of 100 bp
* 55531 56307: contig of 777 bp in length
* 56308 56407: gap of 100 bp
* 56408 57177: contig of 770 bp in length
* 57178 57277: gap of 100 bp
* 57278 57994: contig of 717 bp in length
* 57995 58094: gap of 100 bp
* 58095 58864: contig of 770 bp in length
* 58865 58964: gap of 100 bp
* 58965 59764: contig of 800 bp in length
* 59765 59864: gap of 100 bp
* 59865 60654: contig of 790 bp in length
* 60655 60754: gap of 100 bp
* 60755 61549: contig of 795 bp in length

Query Match 2.5%; Score 47.6; DB 48; Length 80507;
Best Local Similarity 49.2%; Pred. No. 0.25;
Matches 122; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H_NH0387H05

* NOTE: This is a 'working draft' sequence. It currently
* consists of 78 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1287: contig of 1287 bp in length
* 1288 1406: gap of unknown length
* 1407 2639: contig of 1233 bp in length
* 2640 2758: gap of unknown length
* 2759 3949: contig of 1191 bp in length
* 3950 4068: gap of unknown length
* 4069 6120: contig of 2052 bp in length
* 6121 6239: gap of unknown length
* 6240 7854: contig of 1615 bp in length
* 7855 7973: gap of unknown length
* 7974 9166: contig of 1193 bp in length
* 9167 9285: gap of unknown length
* 9286 10469: contig of 1184 bp in length
* 10470 10588: gap of unknown length
* 10589 12886: contig of 2298 bp in length
* 12887 13005: gap of unknown length
* 13006 14222: contig of 1217 bp in length
* 14223 14341: gap of unknown length
* 14342 16064: contig of 1723 bp in length
* 16065 16183: gap of unknown length
* 16184 18245: contig of 2062 bp in length
* 18246 18364: gap of unknown length
* 18365 19779: contig of 1415 bp in length
* 19780 19898: gap of unknown length
* 19899 21787: contig of 1889 bp in length
* 21788 21906: gap of unknown length
* 21907 23682: contig of 1776 bp in length
* 23683 23801: gap of unknown length
* 23802 25816: contig of 2015 bp in length
* 25817 25935: gap of unknown length
* 25936 27544: contig of 1609 bp in length
* 27545 27663: gap of unknown length
* 27664 29356: contig of 1693 bp in length
* 29357 29475: gap of unknown length
* 29476 31588: contig of 2113 bp in length
* 31589 31707: gap of unknown length
* 31708 33985: contig of 2278 bp in length
* 33986 34104: gap of unknown length
* 34105 35708: contig of 1604 bp in length
* 35709 35827: gap of unknown length
* 35828 38639: contig of 2812 bp in length
* 38640 38758: gap of unknown length
* 38759 40502: contig of 1744 bp in length
* 40503 40621: gap of unknown length
* 40622 42742: contig of 2121 bp in length
* 42743 42861: gap of unknown length
* 42862 44999: contig of 2138 bp in length
* 45000 45118: gap of unknown length
* 45119 46962: contig of 1844 bp in length
* 46963 47081: gap of unknown length
* 47082 49731: contig of 2650 bp in length
* 49732 49850: gap of unknown length
* 49851 51830: contig of 1980 bp in length
* 51831 51949: gap of unknown length
* 51950 55175: contig of 3226 bp in length
* 55176 55294: gap of unknown length
* 55295 59415: contig of 4121 bp in length

59416 59534: gap of unknown length
59535 62555: contig of 3021 bp in length
62556 62674: gap of unknown length
62675 68133: contig of 5459 bp in length
68134 68252: gap of unknown length
68253 72424: contig of 4172 bp in length
72425 72543: gap of unknown length
72544 77507: contig of 4964 bp in length
77508 77625: gap of unknown length
83661: contig of 6036 bp in length
83662 83779: gap of unknown length
83780 90116: contig of 6337 bp in length
90117 90234: gap of unknown length
90235 96466: contig of 6232 bp in length
96467 96584: gap of unknown length
96585 102699: contig of 6115 bp in length
102700 102817: gap of unknown length
102818 103977: contig of 1160 bp in length
103978 104095: gap of unknown length
104096 105583: contig of 1488 bp in length
105584 105701: gap of unknown length
105702 106817: contig of 1116 bp in length
106818 106935: gap of unknown length
106936 108385: contig of 1450 bp in length
108386 108503: gap of unknown length
108504 109623: contig of 1120 bp in length
109624 109741: gap of unknown length
109742 110900: contig of 1159 bp in length
110901 111018: gap of unknown length
111019 112324: contig of 1306 bp in length
112325 112442: gap of unknown length
112443 113479: contig of 1037 bp in length
113480 113597: gap of unknown length
113598 114717: contig of 1120 bp in length
114718 114835: gap of unknown length
114836 115955: contig of 1120 bp in length
115956 116073: gap of unknown length
116074 117082: contig of 1009 bp in length
117083 117200: gap of unknown length
117201 118249: contig of 1049 bp in length
118250 118367: gap of unknown length
118368 119584: contig of 1217 bp in length
119585 119702: gap of unknown length
119703 120976: contig of 1274 bp in length
120977 121094: gap of unknown length
121095 122261: contig of 1166 bp in length
122261 122378: gap of unknown length
122379 123710: contig of 1331 bp in length
123710 123827: gap of unknown length
123828 125046: contig of 1219 bp in length
125047 125164: gap of unknown length
125165 126183: contig of 1019 bp in length
126184 126301: gap of unknown length
126302 127831: contig of 1530 bp in length
127832 127949: gap of unknown length
127950 128984: contig of 1035 bp in length
128985 129102: gap of unknown length
129103 130310: contig of 1208 bp in length
130311 130428: gap of unknown length
130429 131514: contig of 1086 bp in length
131515 131632: gap of unknown length
131633 132732: contig of 1100 bp in length
132733 132850: gap of unknown length
132851 134425: contig of 1575 bp in length
134426 134543: gap of unknown length
134544 135606: contig of 1063 bp in length
135607 135724: gap of unknown length
135725 137036: contig of 1312 bp in length
137037 137154: gap of unknown length
137155 138607: contig of 1453 bp in length
138608 138725: gap of unknown length
138726 139899: contig of 1174 bp in length
139900 140017: gap of unknown length


```

; INFORMATION FOR SEQ ID NO: 2:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 1593 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; MOLECULE TYPE: cDNA to mRNA
US-08-524-828-2

```

	Query Match	2.0%;	Score 38.2;	DB 3;	Length 1593;
	Best Local Similarity	50.8%;	Pred. No. 0.15;		
	Matches	91;	Conservative	0;	Mismatches 88; Indels 0; Gaps 0;
QY	915	ccctctccaggcgagccgcgaactggcctgagatgcgcgcctggagagctcccaga	974		
Db	499	CCCACTTGGCGCGGGGCGCTTCGGGCCCGCCCGGGCGCTCAGAGAGGTACGGGG	440		
QY	975	ggcccgccggggtcgcgcgcgacgcgcctcccccacaaacacatctacagcgctctgcccgggc	1034		
Db	439	GTGGCGCGCGGGCGAGAGCGCTTGCGCCCTCAGGCTTCTCTCGGCGCTCCGGGG	380		
QY	1035	gcgccttgagcgacgtctcaggcacagggaggggcccccttcccgcccccgagcgc	1093		
Db	379	CGGCGCTCTCAGACATCTTGCTGCGCGGTTTGAAGCAGACGCCCGCGGCTTGTGGCG	321		

RESULT 6
 US-08-975-114A-2/c
 Sequence 2, Application US/08975114A
 Patent No. 5876714
 GENERAL INFORMATION:
 APPLICANT: Aetsushi NISHIKAWA et al.
 TITLE OF INVENTION: HUMAN GLYCOSYLTRANSFERASE GENE,
 COMPOUNDS AND METHOD FOR INHIBITING CANCEROUS METASTASIS
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
 STREET: 2033 K Street, N.W., Suite 800
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20006
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Wordperfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/975,114A
 FILING DATE: No. 5876714ember 20, 1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/107,173
 FILING DATE: August 17, 1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Lee Cheng
 REGISTRATION NUMBER: 40,949
 REFERENCE/DOCKET NUMBER: 1-F3439DIV
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-721-8200
 TELEFAX: 202-721-8250
 TELEX:
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1593 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cdna to mRNA
 US-08-975-114A-2

	Query Match	2.0%;	Score 38.2;	DB 3;	Length 1593;
	Best Local Similarity	50.8%;	Pred. No. 0.15;		
	Matches 91;	Conservative 0;	Mismatches 88;	Indels 0;	Gaps 0;
QY	915	ccctctccaggcgagccgcgcgaactggccctgagatgcgcgcctggagagctccacga	974		
Db	439	CCCACTTGGCGCGGGCGGCTTCGGCCCCCGTGGCGTCCGGGGCGCTCAGGAGGTACCGGG	440		
QY	975	ggccccgggggtcgcgcgacgcgcgtccccaacacatctacagcgctctgcccgcgcc	1034		
Db	439	GTGGCGCGCGGGCGAGAGCGGTTGGCCCCCTCAGGCTTCTCTCCGGCGCTCCCGGGG	380		
QY	1035	gcgcctctggagcgacctgcgcagcaggagggggcccccdttcccgggcccccgagcgc	1093		
Db	379	CGCGGCTCTCCAGCATCTTTGGTGGCGGGTTTGAAGCAGACGCCGCGGGCTTTGGTGGCG	321		

RESULT 7
 US-08-524-828-1/c
 ; Sequence 1, Application US/08524828
 ; Patent NO. 5874271
 ; GENERAL INFORMATION:
 ; APPLICANT: Atsushi NISHIKAWA et al.
 ; TITLE OF INVENTION: HUMAN GLYCOSYLTRANSFERASE GENE,
 ; TITLE OF INVENTION: COMPOUNDS AND METHOD FOR INHIBITING CANCEROUS METASTASIS
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Weenderoth, Lind & Ponack
 ; STREET: 805 Fifteenth Street, N.W., #700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/524-828
 ; FILING DATE: September 7, 1995
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/107.173
 ; FILING DATE: August 17, 1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: James H. Meadows
 ; REGISTRATION NUMBER: 33,965
 ; REFERENCE/DOCKET NUMBER: JTF/001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-8850
 ; TELEFAX:
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2247 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA to mRNA
 ; US-08-524-828-1

	Query Match	Score 38.2;	DB 3;	Length 2247;
	Best local Similarity 50.6%;			
	Matches 91; Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	915	ccctctcagcgagggccgcgaactgcgcctaggatgcgcgcctcggagacgtccacga	974	
db	667	CCCACTGCGCGGGCGCTCGGCCCCCGCTCCCGGCGCTCAGAGGTACCGGG	608	
QY	975	ggccccgcgggtgcgcgcgaaccgcgctcccaaacacatctacagcgcttcgccgcggc	1034	

Db 607 GTGGCCGCGGCGGAGGCGGTTGGCCCTCAGGCTTCTCTCCGCGCGTCCCGGGG 548
QY 1035 gcgcctctgagcgagcgtcgagggcacagggagggcccccgttccccggcccgagcg 1093
Db 547 GCGGCGCTCTCCAGCATCTTGGTGCGGGGTTTGAAGCAGACGCCGCGGCTTGGTGCGC 489

RESULT 8

US-08-975-114A-1/c
; Sequence 1, Application US/08975114A
; Patent No. 5876714
; GENERAL INFORMATION:
; APPLICANT: Atsushi NISHIKAWA et al.
; TITLE OF INVENTION: HUMAN GLYCOSYLTRANSFERASE GENE,
; TITLE OF INVENTION: COMPOUNDS AND METHOD FOR INHIBITING CANCEROUS METASTASIS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975-114A
; FILING DATE: No. 5876714ember 20, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: 08/107,173
; FILING DATE: August 17, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: 1-F3439DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2247 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-08-975-114A-1

Query Match 2.0%; Score 38.2; DB 3; Length 2247;
Best Local Similarity 50.8%; Pred. No. 0.18;
Matches 91; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 915 ccctctccagggcgccgcgactgaggtgagcgccctggagagctccaga 974
Db 667 CCACATTGCGCGGCGGCGCTCGGCCCGCCCTCGCGGCGCTCAGGAGGTACCGG 608
QY 975 gggcccggtgcgcgcgacgcgtcccaaaacacatctacagcgctgccgcg 1034
Db 607 GTGGCCGCGGCGGAGAGCGGCTTGGCCCTCCTCAGGCTTCTCTCGGCCGTCGCGGG 548
QY 1035 gcgcctggagcgagcgtcgagggcacagggagggcccccgttccccggcccgagcg 1093
Db 547 GCGGCGCTCTCCAGCATCTTGGTGCGGGGTTTGAAGCAGACGCCGCGGCTTGGTGCGC 489

RESULT 9

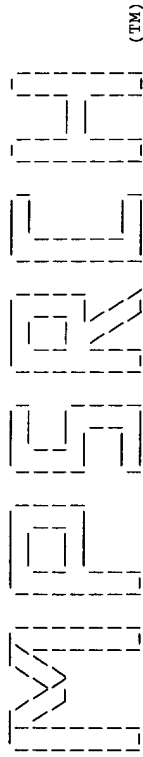
US-07-745-206A-12
; Sequence 12, Application US/07745206A

; Patent No. 5429921
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
; APPLICANT: McCue, Ann
; APPLICANT: Feldman, Daniel
; TITLE OF INVENTION: Human Calcium Channel Compositions and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitch, Even, Tabin & Flannery
; STREET: 135 S. LaSalle
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/745,206A
; FILING DATE: 19910815
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Feder, Scott B
; REFERENCE/DOCKET NUMBER: 51504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-372-7842
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5467 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join(144...3164, 3168..3245, 3249..3386, 3390
; LOCATION: ..3392, 3396..3488, 3495..3539, 3543..3581, 3585
; LOCATION: ..3587, 3591..3626, 3630..3689, 3693..3737, 3744
; LOCATION: ..3746, 3750..4823, 4827..4841, 4845..5006, 5010
; LOCATION: ..5096, 5100..5306, 5310..5366, 5370..5465)
; US-07-745-206A-12
Query Match 2.0%; Score 37.6; DB 1; Length 5467;
Best Local Similarity 51.8%; Pred. No. 0.46;
Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 961 ggagagctccagagggcccggggtcgccgagccggtcccaaaacacatctacag 1020
Db 2765 GGAGAGCGGGGAGCGCGGTGCGCGGAGGCGCGCGCGCGCGGCGCGGCGCGG 2824
QY 1021 gccttgcgcgcgcgctctgtgagcgacgctgcagcgacagggagggcccggtccc 1080
Db 2825 GGAGCGCGGGGCGCGCGCGGAGCGCGGAGCGCGGCGCGGCGCGGCGCGGCGG 2884
QY 1081 gggcccgagagcgctgttcccccccccgccgctgcagggtgtctga 1124
Db 2885 GGCGCGGCGCACCGCGCGGCTCCCGGAGGAGGCGGCGGA 2928
RESULT 10
US-08-311-363-12
; Sequence 12, Application US/08311363
; Patent No. 5876958
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark

Tue Sep 12 15:26:55 2000

Db 1382 CTTGGAGCTGCAGAGCTGAGCGCACAGGGGAACCTCCATTTCAGCGAGCGCGCCCTGG 1323
Qy 1074 cgttcccgcccgagcgcttgcccccgcgcgctgcaggtgtctgaatctccctg 1133
Db 1322 CCTGCCGAGCATCGGTGCGCCTTGGTGCCCTTGTGACCCACTAGGTCTCATCGCACTC 1263
Qy 1134 gctcca 1139
Db 1262 ACTCTA 1257

Search completed: September 12, 2000, 11:49:07
Job time: 598 sec



***** (TM) *****

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Aug 24 10:25:10 2000; MasPar time 9.27 Seconds
Tabular output not generated. 644.932 Million cell updates/sec

Title: >US-09-135-238B-2
Description: (i-390) from US09135238B.pep
Perfect Score: 2830
Sequence: 1 MDRWLWPLYFLPVSGALRIIL.....HQPAAMMEDSDSDYINVPA 390

Scoring table: PAM 150
Gap 11

Searched: 152433 seqs, 15329240 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:6_COMB 4:PCT_COMB 5:backfiles1

Statistics: Mean 34.052; Variance 186.097; scale 0.183

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	254	9.0	771	3	US-08-434-	Sequence 8, Applicatio	3.28e-09
2	247	8.7	746	3	US-08-434-	Sequence 4, Applicatio	1.02e-08
3	245	8.7	757	3	US-08-434-	Sequence 6, Applicatio	1.42e-08
4	246	8.7	769	3	US-08-434-	Sequence 10, Applicati	1.20e-08
5	223	7.9	624	2	US-08-642-	Sequence 22, Applicati	4.91e-07
6	223	7.9	773	3	US-08-434-	Sequence 2, Applicatio	4.91e-07
7	136	4.8	560	1	US-08-336-	Sequence 22, Applicati	3.32e-01
8	129	4.6	201	3	US-08-955-	Sequence 2, Applicatio	9.18e-01
9	123	4.4	46	3	US-08-955-	Sequence 10, Applicati	1.63e+00
10	123	4.3	195	3	US-08-955-	Sequence 4, Applicatio	2.17e+00
11	122	4.3	303	1	US-08-198-	Sequence 2, Applicatio	2.50e+00
12	122	4.3	303	1	US-08-323-	Sequence 2, Applicatio	2.50e+00
13	122	4.3	303	4	PCT-US95-0	Sequence 2, Applicatio	2.50e+00
14	122	4.3	303	4	PCT-US93-0	Sequence 2, Applicatio	2.50e+00
15	118	4.2	319	1	US-08-597-	Sequence 22, Applicati	4.41e+00
16	112	4.0	57	3	US-08-955-	Sequence 5, Applicatio	1.02e+01
17	112	4.0	60	3	US-08-955-	Sequence 23, Applicati	1.55e+01
18	109	3.9	114	1	US-08-211-	Sequence 23, Applicati	1.55e+01
19	109	3.9	114	4	PCT-US93-0	Sequence 23, Applicati	1.55e+01
20	109	3.9	114	1	US-08-111-	Sequence 23, Applicati	1.55e+01
21	110	3.9	134	2	US-08-822-	Sequence 10, Applicati	1.35e+01
22	108	3.8	76	3	US-08-554-	Sequence 21, Applicati	1.78e+01
23	108	3.8	87	3	US-08-554-	Sequence 18, Applicati	1.78e+01

ALIGNMENTS

RESULT 1
ID US-08-434-000A-8 STANDARD; PRT; 771 AA.

XX xxxxxx

DT

XX

DE

XX

Sequence 8, Application US/08434000A

Sequence 8, Application US/08434000A

Patent No. 6046037

GENERAL INFORMATION:

APPLICANT: ANDREW C. HIATT, JULIAN

APPLICANT: K.-C. MA, THOMAS LEHNER

TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon

STREET: 633 West Fifth Street

SUITE: Suite 4700

CITY: Los Angeles

STATE: California

COUNTRY: U.S.A.

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

MEDIUM TYPE: storage

COMPUTER: IBM Compatible

OPERATING SYSTEM: IBM P.C. DOS 5.0

SOFTWARE: Word Perfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/434,000A

FILING DATE: 12/30/94

CLASSIFICATION: 435

PRIOR APPLICATION DATA: including application

PRIOR APPLICATION DATA: described below: 1

APPLICATION NUMBER: 08/367,395

FILING DATE: 12/30/94

ATTORNEY/AGENT INFORMATION:

NAME: Guise, Jeffrey W.

REGISTRATION NUMBER: 34,613

REFERENCE/DOCKET NUMBER: 212/127

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 552-8400

TELEFAX: (619) 552-0159

```
CC      TELEX: 67-3510
CC      TELETYPE: SEQUENCE LISTING
CC      INFORMATION FOR SEQ ID NO: 8:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 771 amino acids
CC      TYPE: amino acid
CC      STRANDEDNESS: single
CC      TOPOLOGY: linear
CC      DESCRIPTION: Mouse Polyimmunoglobulin Receptor
SQ      SEQUENCE 771 AA; 85097 MW; 3010231 CN;

Query Match      9.0%; Score 254; DB 3; Length 771;
Best Local Similarity 36.8%; Pred. No. 3.28e-09;
Matches 43; Conservative 26; Mismatches 42; Indels 6; Gaps 5;

Db 13 FSGVSTKSPFGQEVSSJTEGDSVSIYCPDTSVNRHTRKYWCRO-GASGMCCTLLISSN 71
Qy 10 FLFVSGALRIILPEVKVEGLGGSVTIKC--P-LP-EMHVRILCREMAGSGTCGVIVSTT 65
Db 72 GYLSKEYSGRANLINFENNFTVINIEQLTQDDTGYKGLG-TSNRGLSFDVSLEV 127
Qy 66 NFIKAEYGRVTLKQYPRKNLFLVEVTQLTESDSGVYACGAGMNTDRGKTQKVTLVN 122

RESULT 2
ID US-08-434-000A-4 STANDARD; PRT; 746 AA.
XX
AC xxxxxx
XX
DT
DT
XX
XX
Sequence 4, Application US/08434000A
XX
De Sequence 4, Application US/08434000A
XX
Sequence 4, Application US/08434000A
Patent No. 6046037
CC
CC GENERAL INFORMATION:
CC APPLICANT: ANDREW C. HIATT, JULIAN
CC APPLICANT: K.-C. MA, THOMAS LEHNER
CC TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
CC TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
CC NUMBER OF SEQUENCES: 19
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Lyon & Lyon
CC STREET: 633 West Fifth Street
CC STREET: Suite 4700
CC CITY: Los Angeles
CC STATE: California
CC COUNTRY: U.S.A.
CC ZIP: 90071
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
CC MEDIUM TYPE: storage
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: IBM P.C. DOS 5.0
CC SOFTWARE: Word Perfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/434,000A
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC PRIOR APPLICATION DATA: including application
CC PRIOR APPLICATION DATA: described below: 1
CC APPLICATION NUMBER: 08/367,395
CC FILING DATE: 12/30/94
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Guise, Jeffrey W.
CC REGISTRATION NUMBER: 34,613
CC REFERENCE/DOCKET NUMBER: 212/127
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 552-8400
CC TELEFAX: (619) 552-0159
CC TELEX: 67-3510
CC TELETYPE: SEQUENCE LISTING
CC
```

```
CC      INFORMATION FOR SEQ ID NO: 4:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 746 amino acids
CC      TYPE: amino acid
CC      STRANDEDNESS: single
CC      TOPOLOGY: linear
CC      DESCRIPTION: Human Polyimmunoglobulin Receptor
SQ      SEQUENCE 746 AA; 81378 MW; 2795658 CN;

Query Match      8.7%; Score 247; DB 3; Length 746;
Best Local Similarity 37.7%; Pred. No. 1.02e-08;
Matches 40; Conservative 27; Mismatches 32; Indels 7; Gaps 5;

Db 7 PE-EVNSVGNVSITCYPTSVNRHTRKYWCRO-GARGGCTITLSSSEGVSSKYAGRA 64
Qy 21 PEVKVEGLGGSVTIKC---PLP-EMHVRILCREMAGSGTCGVIVSTTNRIKAEYGRV 76
Db 65 NLNFPENFTFVNIAQLSDSDSGRYKGLGINS-RGLSFDVSLEV 109
Qy 77 TLKQYPRKNLFLVEVTQLTESDSGVYACGAGMNTDRGKTQKVTLVN 122

RESULT 3
ID US-08-434-000A-6 STANDARD; PRT; 757 AA.
XX
AC xxxxxx
XX
DT
DT
XX
XX
Sequence 6, Application US/08434000A
XX
Sequence 6, Application US/08434000A
Patent No. 6046037
CC
CC GENERAL INFORMATION:
CC APPLICANT: ANDREW C. HIATT, JULIAN
CC APPLICANT: K.-C. MA, THOMAS LEHNER
CC TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
CC TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
CC NUMBER OF SEQUENCES: 19
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Lyon & Lyon
CC STREET: 633 West Fifth Street
CC STREET: Suite 4700
CC CITY: Los Angeles
CC STATE: California
CC COUNTRY: U.S.A.
CC ZIP: 90071
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
CC MEDIUM TYPE: storage
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: IBM P.C. DOS 5.0
CC SOFTWARE: Word Perfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/434,000A
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC PRIOR APPLICATION DATA: including application
CC PRIOR APPLICATION DATA: described below: 1
CC APPLICATION NUMBER: 08/367,395
CC FILING DATE: 12/30/94
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Guise, Jeffrey W.
CC REGISTRATION NUMBER: 34,613
CC REFERENCE/DOCKET NUMBER: 212/127
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 552-8400
CC TELEFAX: (619) 552-0159
CC TELEX: 67-3510
CC TELETYPE: SEQUENCE LISTING
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC
```

CC LENGTH: 757 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC TOPOLOGY: DESCRIPTION: Bovine Polymunoglobulin Receptor
SQ SEQUENCE 757 AA; 82450 MW; 2942196 CN;

Query Match 8.7%; Score 245; DB 3; Length 757;
Best Local Similarity 36.7%; Pred. No. 1.42e-08;
Matches 40; Conservative 25; Mismatches 37; Indels 7; Gaps 5;

Db 25 PE-EVSVSVEGRSVSKYPPPTSVNRHTRKYWCRO-GAQRCTLTLSSEGYSDDYVGRA 82
QY 21 PEVKVEGELGGSVTIKC---PLP-EMHVIYLCREMAGSGTCGTVTSTTFIKAEYKGRV 76
Db 83 NLTNPESGTFVVDISHLTHDKSGYKGLGIS-RGLNFDVSLVSQD 130
QY 77 TLKQIPRNLFLVEVTQLTESDSGVYACGAGMNTDRGKTQKVTINVHSE 125

RESULT 4
ID US-08-434-000A-10 STANDARD; PRT; 769 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 10, Application US/08434000A
Sequence 10, Application US/08434000A
Patent No. 6045037
GENERAL INFORMATION:
APPLICANT: ANDREW C. HIATT, JULIAN
APPLICANT: K.-C. MA, THOMAS LEHNER
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: LYON & LYON
STREET: 633 West Fifth Street
CITY: Suite 4700
STATE: Los Angeles
COUNTRY: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,000A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application 1
PRIOR APPLICATION DATA: described below:
APPLICATION NUMBER: 08/367,395
FILING DATE: 12/30/94
ATTORNEY/AGENT INFORMATION:
NAME: Guise, Jeffrey W.
REGISTRATION NUMBER: 34,613
REFERENCE/DOCKET NUMBER: 212/127
TELEPHONE: (619) 552-8400
TELEFAX: (619) 552-0159
TELEX: 67-3510
TELEX: SEQUENCE LISTING
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid

CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC TOPOLOGY: DESCRIPTION: Rat Polymunoglobulin Receptor
SQ SEQUENCE 769 AA; 84798 MW; 2960864 CN;

Query Match 8.7%; Score 246; DB 3; Length 769;
Best Local Similarity 35.0%; Pred. No. 1.20e-08;
Matches 41; Conservative 30; Mismatches 40; Indels 6; Gaps 5;

Db 13 FSGVSTOSPFGQDYSVSIENSVSITCYPPDPISVNRHTRKYWCRO-GANGYCATLISSN 71
QY 10 FLPVSGALRILPEVKVEGELGGSVTIKC--P-LP-EMHVIYLCREMAGSGTCGTVWST 65
Db 72 GYLSKEYSGRASLINEFENSTFVINIAHLTQEDTGSYKGLG-TTNRGLFFDVSLV 127
QY 66 NFIAEKYKGRVTLKQIPRNLFLVEVTQLTESDSGVYACGAGMNTDRGKTQKVTINLV 122

RESULT 5
ID US-08-642-406A-22 STANDARD; PRT; 624 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 22, Application US/08642406A
Sequence 22, Application US/08642406A
Patent No. 5959177
GENERAL INFORMATION:
APPLICANT: Hein, Mich B.
APPLICANT: Hiatt, Andrew C.
APPLICANT: Ma, Julian K.C.
TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED
TITLE OF INVENTION: SECRETORY ANTIBODIES
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
STREET: 10666 No. 5959177th Torrey Pines Road, TPC-8
CITY: La Jolla
STATE: California
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIP Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,406A
FILING DATE: 03-MAY-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/591,823
FILING DATE: 02-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/427,765
FILING DATE: 27-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Logan, April C.
REGISTRATION NUMBER: 33,950
REFERENCE/DOCKET NUMBER: 184.2
TELEPHONE: (619) 554-2937
TELEFAX: (619) 554-6312
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 624 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 624 AA; 68416 MW; 2044676 CN;

CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/323,531
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/936,784
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: White, John P
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)977-9550
CC TELEFAX: (212)664-0525
CC TELEX: 422523
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 303 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 303 AA; 33465 MW; 451890 CN;

Query Match 4.3%; Score 122; DB 1; Length 303;
Best Local Similarity 32.4%; Pred. No. 2.50e+00;
Matches 22; Conservative 19; Mismatches 23; Indels 4; Gaps 4;

Db 224 PARLEIPAPGRPLPRRP-SEGGM-RAPRRSRAPAAARSTAAATPPRPGDPAPAAAR 281
Qy 290 AVRMRALSSORPGSRPRSONNIYSACPRRAGADAA-GTGEAPVPG-PGAPLPAPL 347
Db 282 RAGDVTWM 289
Qy 348 QVSESPWL 355

RESULT 13
ID PCT-US95-02087-2 STANDARD; PRT; 303 AA.

XX xxxxxx

Sequence 2, Application PC/TUS9502087

Sequence 2, Application PC/TUS9502087
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D
TITLE OF INVENTION: Recombinant Equine Herpesviruses
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02087
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/198,094
FILING DATE: February 17, 1994
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678

CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)278-0400
CC TELEFAX: (212)391-0525
CC TELEX: 422523
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 303 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 303 AA; 33465 MW; 451890 CN;

Query Match 4.3%; Score 122; DB 4; Length 303;
Best Local Similarity 32.4%; Pred. No. 2.50e+00;
Matches 22; Conservative 19; Mismatches 23; Indels 4; Gaps 4;

Db 224 PARLEIPAPGRPLPRRP-SEGGM-RAPRRSRAPAAARSTAAATPPRPGDPAPAAAR 281
Qy 290 AVRMRALSSORPGSRPRSONNIYSACPRRAGADAA-GTGEAPVPG-PGAPLPAPL 347
Db 282 RAGDVTWM 289
Qy 348 QVSESPWL 355

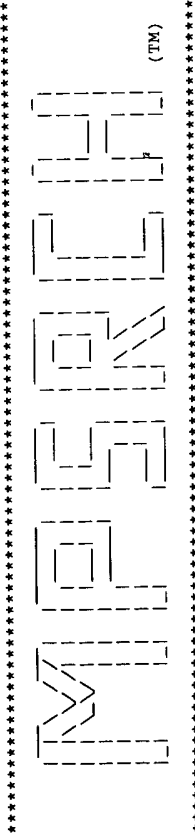
RESULT 14
ID PCT-US93-07424-2 STANDARD; PRT; 303 AA.

XX xxxxxx

Sequence 2, Application PC/TUS9307424

Sequence 2, Application PC/TUS9307424
GENERAL INFORMATION:
APPLICANT: Cochran Ph.D., Mark D
TITLE OF INVENTION: Recombinant Equine Herpesvirus
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07424
FILING DATE: 19930806
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)977-9550
TELEFAX: (212)664-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 303 AA; 33465 MW; 451890 CN;

Query Match 4.3%; Score 122; DB 4; Length 303;
Best Local Similarity 32.4%; Pred. No. 2.50e+00;
Matches 22; Conservative 19; Mismatches 23; Indels 4; Gaps 4;



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd
MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Aug 24 10:22:24 2000; MasPar time 13.89 Seconds
Tabular output not generated.

Title: >US-09-135-238B-2
Description: (1-390) from US09135238B.pep
Perfect Score: 2830
Sequence: 1 MDRWLWFLYFLVPVSGALRIL.....HQPAAEMEDSDDDYINVPA 390

Scoring table: PAM 150
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 50.211; Variance 94.729; scale 0.530

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	254	9.0	771	1	P1GR_MOUSE POLYMERIC-IMMUNOGLOBUL	1.08e-25
2	252	8.9	764	1	P1GR_HUMAN POLYMERIC-IMMUNOGLOBUL	2.54e-25
3	246	8.7	769	1	P1GR_RAT POLYMERIC-IMMUNOGLOBUL	3.32e-24
4	244	8.6	757	1	P1GR_BOVIN POLYMERIC-IMMUNOGLOBUL	7.79e-24
5	223	7.9	773	1	P1GR_RABIT POLYMERIC-IMMUNOGLOBUL	5.43e-20
6	188	6.6	102	1	P1GR_PIG POLYMERIC-IMMUNOGLOBUL	8.21e-14
7	140	4.9	224	1	CM35_HUMAN CMRF35 ANTIGEN PRECURS	5.46e-06
8	126	4.5	3321	1	KEND_HUMAN KENDRIN (KIAA0402).	6.49e-04
9	119	4.2	303	1	US02_HSVK US1 PROTEIN	6.35e-03
10	118	4.2	319	1	A33_HUMAN CELL SURFACE A33 ANTIG	8.74e-03
11	115	4.1	116	1	SP22_RAT NUCLEAR TRANSITION PRO	2.29e-02
12	115	4.1	179	1	VG59_HSVB HYPOTHETICAL GENE 59 P	2.25e-02
13	117	4.1	567	1	CVDC_BAGSU TRANSPORT ATP-BINDING	1.20e-02
14	114	4.0	117	1	SP22_MOUSE NUCLEAR TRANSITION PRO	3.08e-02
15	114	4.0	146	1	HY21_HUMAN IG HEAVY CHAIN V-II RE	3.08e-02
16	113	4.0	863	1	MCM4_HUMAN DNA REPLICATION LICENS	4.20e-02
17	109	3.9	502	1	VP22_HPV25 REGULATORY PROTEIN E2.	1.43e-01
18	108	3.8	252	1	VPHE_NPVAC POLYHEDRAL ENVELOPE PR	1.93e-01
19	108	3.8	352	1	P2Y7_HUMAN P2Y PURINOCEPTOR 7 (P2	1.93e-01
20	108	3.8	358	1	VCOM_ADE40 MINOR CORE PROTEIN (PR	1.93e-01
21	104	3.7	208	1	YSX2_CAEBL HYPOTHETICAL 24.0 KDA	6.28e-01
22	104	3.7	347	1	CSP_PLA8A CIRCUMSPOROZOITE PROTE	6.28e-01
23	104	3.7	483	1	VE2_HPV14 REGULATORY PROTEIN E2.	6.28e-01

RESULT ID	P1GR_MOUSE	STANDARD;	PRT;	771 AA.
AC	O70570;			
DT	15-FEB-2000 (Rel. 39, Created)			
DT	15-FEB-2000 (Rel. 39, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (POLY-IG RECEPTOR) (P1GR)			
DE	[CONTAINS: SECRETORY COMPONENT].			
GN	P1GR.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6 X CBA; TISSUE=LIVER;			
RX	MEDLINE; 95138517.			
RA	Piskurich J.F., Blanchard M.H., Youngman K.R., France J.A.,			
RA	Kaetzel C.S.;			
RT	"Molecular cloning of the mouse polymeric Ig receptor. Functional			
RT	regions of the molecule are conserved among five mammalian species."			
RL	J. Immunol. 154:1735-1747(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129/SVJ;			
RX	MEDLINE; 98072444.			
RA	Martin M.G., Gutierrez E.M., Lam J.T., Li T.W.H., Wang J.;			
RT	"Genomic cloning and structural analysis of the murine polymeric			
RT	receptor (p1gr) gene and promoter region."			
RL	Gene 201:189-197(1997).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129; TISSUE=LIVER;			
RA	de Groot N., Vollebregt E., Lee S.H., Verbeet M.P., de Boer H.A.;			
RT	"Molecular cloning and exon-intron organization of the gene encoding			
RT	the murine polymeric immunoglobulin receptor."			
RL	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE			
CC	BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN			
CC	TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.			
CC	DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE			
CC	EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE			
CC	TRANSMEMBRANE SEGMENT (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.			
CC	-1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U06431; AAC67440.1; -;
 DR EMBL; U83434; AAC53585.1; -;
 DR EMBL; U83427; AAC53585.1; JOINED.
 DR EMBL; U83428; AAC53585.1; JOINED.
 DR EMBL; U83429; AAC53585.1; JOINED.
 DR EMBL; U83430; AAC53585.1; JOINED.
 DR EMBL; U83431; AAC53585.1; JOINED.
 DR EMBL; U83432; AAC53585.1; JOINED.
 DR EMBL; U83433; AAC53585.1; JOINED.
 DR EMBL; Y16524; CAA76272.1; -;
 DR EMBL; Y16525; CAA76272.1; JOINED.
 DR EMBL; Y16526; CAA76272.1; JOINED.
 DR EMBL; Y16527; CAA76272.1; JOINED.
 DR EMBL; Y16528; CAA76272.1; JOINED.
 DR EMBL; Y16529; CAA76272.1; JOINED.
 DR EMBL; Y16530; CAA76272.1; JOINED.
 DR EMBL; Y16531; CAA76272.1; JOINED.
 DR EMBL; Y16532; CAA76272.1; JOINED.
 DR MGD; MGI:23029; PIGR.
 DR PFAM; PF00047; ig.5.
 KW Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein; Signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 771 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.
 FT CHAIN 19 611 SECRETORY COMPONENT.
 FT DOMAIN 19 645 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 646 668 POTENTIAL.
 FT DOMAIN 669 771 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 33 117 IG-LIKE V-TYPE DOMAIN 1.
 FT DOMAIN 145 227 IG-LIKE V-TYPE DOMAIN 2.
 FT DOMAIN 250 331 IG-LIKE V-TYPE DOMAIN 3.
 FT DOMAIN 363 447 IG-LIKE V-TYPE DOMAIN 4.
 FT DOMAIN 477 553 IG-LIKE V-TYPE DOMAIN 5.
 FT DISULFID 40 110 POTENTIAL.
 FT DISULFID 152 220 POTENTIAL.
 FT DISULFID 257 324 POTENTIAL.
 FT DISULFID 370 440 POTENTIAL.
 FT DISULFID 484 546 POTENTIAL.
 FT CARBOHYD 90 90 POTENTIAL.
 FT CARBOHYD 147 147 POTENTIAL.
 FT CARBOHYD 170 170 POTENTIAL.
 FT CARBOHYD 206 206 POTENTIAL.
 FT CARBOHYD 420 420 POTENTIAL.
 FT CARBOHYD 471 471 POTENTIAL.
 FT CONFLICT 159 159 A -> V (IN REF. 1).
 FT CONFLICT 396 396 V -> A (IN REF. 1).
 FT CONFLICT 620 620 G -> R (IN REF. 1).
 SQ SEQUENCE 771 AA: 84998 MW: 78C81302EC710730 CRC64;
 Query Match 9.0%; Score 254; DB 1; Length 771;
 Best Local Similarity 36.8%; Pred. No. 1.08e-25;
 Matches 43; Conservative 26; Mismatches 42; Indels 6; Gaps 5;
 Db 13 FSGVSTKSPFGQEVSSIEGDSVSTCYPTDTSVNRHTRKWKCRQ-GASGMCTTLISSN 71
 QY 10 FLFVSGALRILPEVKVEGLGGSTKIC--P-LP-EMHVIYLCREMAGSGTCTGVSTT 65
 Db 72 GYLKSEYSGRANLINFNNFTVNIETQDDTGSYKCGLG-TSNRGLSFDVSLEV 127
 QY 66 NFPAKAYKGRVTLKQYPRKNFLFVLTQTESDSGVACGAMNTDRGKTQKVLNV 122
 RESULT 2
 ID PIGR_HUMAN STANDARD; PRT; 764 AA.
 AC P01833;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-OCT-1986 (Rel. 34, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)

DE DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (POLY-IG RECEPTOR) (PIGR)
 GN [CONTAINS: SECRETORY COMPONENT].
 OS PIGR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 92039621.
 RA Krajci P., Grzeschik K.H., Geurts van Kessel A.H., Olaisen B.,
 RA Brandtzaeg P.;
 RT "The human transmembrane secretory component (poly-Ig receptor):
 RT molecular cloning, restriction fragment length polymorphism and
 RT chromosomal sublocalization.";
 RL Hum. Genet. 87:642-648(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 92387236.
 RA Krajci P., Kvale D., Tasken K., Brandtzaeg P.;
 RT "Molecular cloning and exon-intron mapping of the gene encoding human
 RT transmembrane secretory component (the poly-Ig receptor).";
 RL Eur. J. Immunol. 22:2309-2315(1992).
 RN [3]
 RP SEQUENCE OF 72-764 FROM N.A.
 RX MEDLINE; 89149795.
 RA Krajci P., Solberg R., Sandberg M., Oyen O., Jahnsen T.,
 RA Brandtzaeg P.;
 RT "Molecular cloning of the human transmembrane secretory component
 RT (poly-Ig receptor) and its mRNA expression in human tissues.";
 RL Biochem. Biophys. Res. Commun. 158:783-789(1989).
 RN [4]
 RP SEQUENCE OF 19-577, DISULFIDE BONDS, AND CARBOHYDRATE-BINDING SITES.
 RX MEDLINE; 85128981.
 RA Eifert H., Quentin E., Decker J., Hillemeir S., Hufschmidt M.,
 RA Klingmüller D., Weber M.H., Hilschmann N.;
 RT "The primary structure of human free secretory component and the
 RT arrangement of disulfide bonds.";
 RL Hoppe-Seyler's Z. Physiol. Chem. 365:1489-1495(1984).
 RN [5]
 RP SEQUENCE OF 19-577.
 RX MEDLINE; 91315750.
 RA Eifert H., Quentin E., Wiederhold M., Hillemeir S., Decker J.,
 RA Weber M., Hilschmann N.;
 RT "Determination of the molecular structure of the human free secretory
 RT component.";
 RL Biol. Chem. Hoppe-Seyler 372:119-128(1991).
 RN [6]
 RP SEQUENCE OF 118-138; 212-230; 232-268; 273-288 AND 578-603.
 RX MEDLINE; 97379357.
 RA Hughes G.J., Frutiger S., Savoy L.-A., Reason A.J., Morris H.R.,
 RA Jaton J.-C.;
 RT "Human free secretory component is composed of the first 595 amino
 RT acid residues of the polymeric immunoglobulin receptor.";
 RL FEBS Lett. 410:443-446(1997).
 CC -!- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
 CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN
 CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.
 CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
 CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE
 CC TRANSMEMBRANE SEGMENT.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.
 CC -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S62403; AAB20203.1; -;
 DR EMBL; S43449; AAB23176.1; -;

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: X00412; CAA25118.1; -.
PIR: A02111; QRRBG.
PIR: A28077; A28077.
PFAM: PF00047; ig; 5.
Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein; Signal; Polymorphism.

FT SIGNAL 1 18 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.
FT CHAIN 19 773 SECRETORY COMPONENT.
FT CHAIN 19 615 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 19 647 POTENTIAL.
FT TRANSMEM 648 670 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 671 773 IG-LIKE V-TYPE DOMAIN 1.
FT DOMAIN 30 136 IG-LIKE V-TYPE DOMAIN 2.
FT DOMAIN 243 243 IG-LIKE V-TYPE DOMAIN 3.
FT DOMAIN 244 350 IG-LIKE V-TYPE DOMAIN 4.
FT DOMAIN 351 456 IG-LIKE V-TYPE DOMAIN 5.
FT DOMAIN 457 558 POTENTIAL.
FT DISULFID 46 115 POTENTIAL.
FT DISULFID 155 225 POTENTIAL.
FT DISULFID 260 324 POTENTIAL.
FT DISULFID 369 438 POTENTIAL.
FT DISULFID 478 538 POTENTIAL.
FT CARBOHYD 108 108 IN ALLOTYPIC T61.
FT CARBOHYD 418 418 IN ALLOTYPIC T62 (PARTIAL) AND T63.
FT VARIANT 88 88 K -> N (IN ALLOTYPIC T61).
FT VARIANT 94 94 D -> E (IN ALLOTYPIC T61).
FT VARIANT 101 108 TVDOLTON -> YLNLSOS (IN ALLOTYPIC T61).
FT VARIANT 110 110 S -> T (IN ALLOTYPIC T63).
SQ SEQUENCE 773 AA; 83886 MW; DF2G4D2F1193G65 CRC64;

Query Match 7.9%; Score 223; DB 1; Length 773;

Best Local Similarity 45.5%; Pred. No. 5.43e-20; Mismatches 11; Indels 7; Gaps 5;

Matches 40; Conservative 11; Mismatches 30; Indels 7; Gaps 5;
Db 39 GDSVITCYPTTSVTRHSRKFRCRE-EESGRVCVTLAST-GYTSQESYGRGKLTDFPDG 96
Qy 30 GGSVTKC-PLPEM-HVRIYLCREMGSGTGVSTVTFINFAEYKGRVTLKQYPRKN 85
Db 97 EFVTVTDLQNDSGSKGCVGVN-GRG 123
Qy 86 LFLVEVTQLTESDSGVYACGAGMNTDRG 113

RESULT 6 STANDARD; PRT; 102 AA.

ID PIGR_PIG Q29244;
AC Q29244;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR (POLY-IG RECEPTOR) (PIGR)
DE (FRAGMENT).
OS Sus scrofa (Pig).
GN PIGR.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SMALL INTESTINE;
RA Winteroe A.K., Fredholm M., Davies W.;
RT "Evaluation and characterization of a porcine small intestine cDNA library: analysis of 839 clones.";
RL Mamm. Genome 7:509-517(1996)
CC -!- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.

-!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: F14851; CAA23294.1; -.
Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein.

FT NON_TER 1 102
FT NON_TER 102 102
SQ SEQUENCE 102 AA; 11205 MW; 82C915264B1508E8 CRC64;

Query Match 6.6%; Score 188; DB 1; Length 102;
Best Local Similarity 38.3%; Pred. No. 8.21e-14;
Matches 36; Conservative 19; Mismatches 31; Indels 8; Gaps 8;

Db 2 LOVLKPEPELYDGLRGSVTFDCALQEMANVAKFLCOLKNGK-TCNVVINTLG-KKAQD 59
Qy 17 LRIL-PEVK-VEGELGSGVTIKCPL-PEM-HVRIYLCREMGSGTGVSTVTFNFIKA-E 71

Db 60 PEGRILLTP-KENSHFSVHITGLRKEDAGHYLCG 92

Qy 72 YKGRVTLKQYPRKNLFLVEVTQLTESDSGVYACG 105

RESULT 7

ID CM35_HUMAN STANDARD; PRT; 224 AA.
AC Q08708;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CMRF35 ANTIGEN PRECURSOR.
GN CMRF35.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92249405.
RA Jackson D.G., Hart D.N.J., Starling G., Bell J.I.;
RT "Molecular cloning of a novel member of the immunoglobulin gene superfamily homologous to the polymorphic immunoglobulin receptor.";
RL Eur. J. Immunol. 22:1157-1163(1992).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -!- TISSUE SPECIFICITY: PRESENT ON THE SURFACE OF MONOCYTES, NEUTROPHILS, A PROPORTION OF PERIPHERAL BLOOD T AND B LYMPHOCYTES AND LYMPHOCYTIC CELL LINES.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: X66171; CAA46948.1; -.
DR PFAM: PF00047; ig; 1.
DR Antigen; Transmembrane; Glycoprotein; Signal.
KW SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 224 CMRF35 ANTIGEN.
FT DOMAIN 21 183 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 184 204 POTENTIAL.
FT DOMAIN 205 224 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 36 117 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 123 175 PRO-RICH.
FT DISULFID 43 110 POTENTIAL.
FT DISULFID 57 65

[W][A][T][E][R][S][E][R][E][S][E][R][E]

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Aug 24 10:22:59 2000; Maspar time 32.75 Seconds
Tabular output not generated. 825.707 Million cell updates/sec

Title: >US-09-135-238B-2
Description: (1-390) from US09135238B.pep
Perfect Score: 1
Sequence: 1 MDRWLWPLFLPVSGALRIL.....HOPAAAMEDSDSDDYINVPA 390

Scoring table: PAM 150
Gap 11
Searched: 225878 seqs, 69334122 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrmlb12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 49.245; Variance 97.409; scale 0.506
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match %	Description	Pred. No.
1	2815	99.5	ANTI-FAS-INDUCED APOPT	0.00e+00
2	149	5.3	CMRF-35-H9.	1.01e-06
3	129	4.6	NUCLEAR TRANSITION PRO	8.53e-04
4	129	4.6	NUCLEAR TRANSITION PRO	8.53e-04
5	129	4.6	L2385.5 PROTEIN.	8.53e-04
6	126	4.5	KIAA0402 (FRAGMENT).	2.24e-03
7	126	4.5	KENDRIN.	2.24e-03
8	122	4.3	HYPOTHETICAL 23.8 KD P	7.99e-03
9	119	4.2	BCDNA.GH10711.	2.04e-02
10	115	4.1	KIAA0592 PROTEIN (FRAG	6.95e-02
11	112	4.0	EUS1 PROTEIN.	1.71e-01
12	113	4.0	LEUKOTRIENE B4 RECEPTO	1.27e-01
13	113	4.0	KOZAL1.1 PROTEIN.	1.27e-01
14	113	4.0	HYPOTHETICAL 53.3 KD P	1.27e-01
15	114	4.0	HYPOTHETICAL 91.0 KD P	9.40e-02
16	109	3.9	F08H9.2 PROTEIN.	4.15e-01
17	110	3.9	AMBIENTAL PH REGULATOR	3.10e-01
18	111	3.9	SIALIDASE (EC 3.2.1.18	2.31e-01
19	111	3.9	KIAA0973 PROTEIN (FRAG	2.31e-01
20	109	3.9	FATTY ACID SYNTHASE.	4.15e-01

21	108	3.8	289 10	Q9XF31	HYPOTHETICAL 31.7 KD P	5.56e-01
22	108	3.8	395 1	Q9Y918	395AA LONG HYPOTHETICA	5.56e-01
23	107	3.8	654 5	Q17982	SIMILARITY TO ERBB-3 R	7.43e-01
24	108	3.8	699 10	O81765	EXTENSIN-LIKE PROTEIN.	5.56e-01
25	107	3.8	944 4	Q14163	KIAA0150 PROTEIN (FRAG	7.43e-01
26	107	3.8	1353 3	O74162	ICHI.	7.43e-01
27	108	3.8	3247 14	Q65553	UL36.	5.56e-01
28	105	3.7	57 14	O55723	PUTATIVE SMALL BASIC P	1.32e+00
29	104	3.7	178 2	O9XAB5	PUTATIVE INTEGRAL MEMB	1.75e+00
30	106	3.7	416 10	O9XE67	HYPOTHETICAL 46.5 KD P	9.91e-01
31	105	3.7	508 10	O49444	COPPER-BINDING PROTEIN	1.32e+00
32	106	3.7	539 2	O83771	HYPOTHETICAL 63.1 KD P	9.91e-01
33	105	3.7	766 10	Q39177	PEARL 4.	1.32e+00
34	104	3.7	776 11	Q92143	SEMAPHORIN W.	1.75e+00
35	104	3.7	815 5	O97214	L4830.10 PROTEIN.	1.75e+00
36	106	3.7	884 4	O00302	ERPROT 213-21.	9.91e-01
37	104	3.7	886 13	P70049	ORIGIN RECOGNITION COM	1.75e+00
38	106	3.7	897 11	O70495	PLENTY-OF-PROLINES-101	9.91e-01
39	103	3.6	126 3	Q12444	ORF YOR309C.	2.32e+00
40	103	3.6	279 2	O86663	PUTATIVE INTEGRAL MEMB	2.32e+00
41	102	3.6	662 5	Q21139	SIMILAR TO DROSOPHILA	3.07e+00
42	102	3.6	1179 4	O14120	DBP-5 NUCLEAR PROTEIN.	3.07e+00
43	102	3.6	1280 4	O75158	KIAA0670 PROTEIN (FRAG	3.07e+00
44	103	3.6	1336 14	O73550	POLYPROTEIN.	2.32e+00
45	103	3.6	1792 6	O46385	SUPERVILLIN.	2.32e+00

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	390 AA.
ID	O60667			
AC	O60667			
DT	01-AUG-1998 (Tremblrel. 07, Created)			
DT	01-AUG-1998 (Tremblrel. 07, Last sequence update)			
DT	01-MAY-1999 (Tremblrel. 10, Last annotation update)			
DE	ANTI-FAS-INDUCED APOPTOSIS.			
GN	TOSO.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Hominidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE; 98246048.			
RA	HITOSHI Y., LORENS J., KITADA S.I., FISHER J., LABARGE M., RING H.Z.,			
RA	FRANCKE U., REED J.C., KINOSHITA S., NOLAN G.P.;			
RT	"Toso, a cell surface, specific regulator of Fas-induced apoptosis in			
RT	T cells.";			
RL	Immunity 8:461-471(1998).			
DR	EMBL; AF057557; AAC18830.1; -.			
DR	PFAM; PF00047; ig; 1.			
SQ	SEQUENCE 390 AA; 43146 MW; 15D1B5D8 CRC32;			

Query Match	99.5%;	Score 2815;	DB 4;	Length 390;
Best Local Similarity	99.7%;	Pred. No. 0.00e+00;		
Matches	389;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;
Db	1	MDFWLWPLFLPVSGALRILPEVKVEGELGGSVTKICPLPEMHVRIYLCREMGSGT	CGT 60	
Qy	1	MDRWLWPLFLPVSGALRILPEVKVEGELGGSVTKICPLPEMHVRIYLCREMGSGT	CGT 60	
Db	61	VVSTTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTSDSGVYACAGMNTDRGKTQVTL	120	
Qy	61	VVSTTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTSDSGVYACAGMNTDRGKTQVTL	120	
Db	121	NHSEPEPWEQPMETPKWFLYLFQMPAYASSKFVTRVTPPAQKGKVPVHHSSP	180	
Qy	121	NHSEPEPWEQPMETPKWFLYLFQMPAYASSKFVTRVTPPAQKGKVPVHHSSP	180	
Db	181	TQIITHRPVSRASSVAGDKPRTFLPSTTASISALEGLLKPTQPSYNNHTLHRQALD	240	
Qy	181	TQIITHRPVSRASSVAGDKPRTFLPSTTASISALEGLLKPTQPSYNNHTLHRQALD	240	
Db	241	YGSQSGREGGFHILPTITLGLLALLGLVVKRAVERRKALSRARLAVRMRALSSQ	300	

RT	²³⁸ K
RL	Su


```

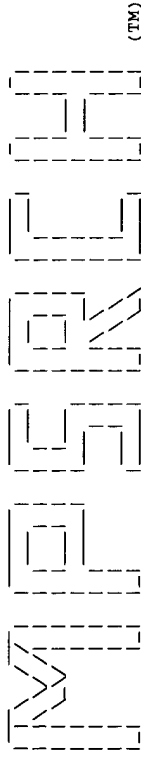
Db      165 RTLICAPNYPNKHKVPHLLFENAITG-FLLPFLAVNVASVDIGRRRLQARRFRSRR-TGR 222
       1 : | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY     237 RALDYGSG-SRGEGGFHILPTILGLIFLALLGLGVK-RAVERR-KALS-RRARLAVR 292
       1 : | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      223 LVVL 226
       :   |
QY     293 MRAL 296
       :   |

RESULT 13
ID      P90902    PRELIMINARY;          PRT;        476 AA.
AC      Q90902;
DT      01-MAY-1997 (TEMBLrel. 03, Created)
DT      01-MAY-1997 (TEMBLrel. 03, Last sequence update)
DT      01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE      K02A11.1 PROTEIN.
DE      K02A11.1 GN
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiida;
OC      Rhabditiina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN      [1]
RA      SWYE R.;
RL      Submitted (JUN-1996) to the EMBL/GenBank/DBDJ databases.
RN      [2]
SEQUENCE FROM N.A.
RX      MEDLINE; 94150718.
RA      WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA      BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA      CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA      GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA      JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA      LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA      PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA      SALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA      THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA      WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT      "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT      elegans.";
RL      Nature 368:32-38(1994).
DR      EMBL; Z75544; CAA99881.1; -.
DR      HSSP; P55273; lBD8.
DR      PFAM; PF00023; ank; 4.
SQ      SEQUENCE 476 AA; 53687 MW; 2F755235 CRC32;

Query Match           4.0%; Score 113; DB 5; Length 476;
Best Local Similarity 35.2%; Pred. No. 1.27e-01;
Matches              19; Conservative 13; Mismatches 20; Indels 2; Gaps

Db      6 LIAELPIERTAQETALADRRLRLQLR-RN-DERERALPPPRQRRLHFPA 57
       1 : | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY     264 LLALLGLVWKRVERRKALSRARRLAVRMALESSORPGSRPSRQNNIYS 317
       1 : | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 14
ID      Q9Y4M4    PRELIMINARY;          PRT;        491 AA.
AC      Q9Y4M4;
DT      01-NOV-1999 (TEMBLrel. 12, Created)
DT      01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DT      01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE      HYPOTHEtical 53.3 KD PROTEIN (FRAGMENT).
DE      DKFPZP4340171.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC      Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN      [1]
RA      SEQUENCE FROM N.A.
RC      TISSUE=TESTIS;
RA      POUSTKA A., KLEIN M., MEWES H.W., CASSENHUBER J., WIEMANN S.;
RL      Submitted (JUN-1999) to the EMBL/GenBank/DBDJ databases.
DR      EMBL; AL080183; CAB45765.1; -.
KW      Hypothetical protein.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Aug 24 10:23:55 2000; MasPar time 21.42 Seconds
859.002 Million cell updates/sec
Tubular output not generated.

Title: >US-09-135-238B-2
Description: (1-390) from US09135238B.pep
Perfect Score: 2830
Sequence: 1 MDRWLWLPVLPVSGALRL.....HQPAMMEDSDSDYINVPA 390

Scoring table:
PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir64
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 49.586; Variance 110.938; scale 0.447

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	ID	Description	Pred. No.
1	252	8.9	764	1	QRUGS secretory component p	6.59e-21
2	245	8.7	757	1	S48841 secretory component p	8.34e-20
3	246	8.7	769	1	QRRTGS secretory component p	5.81e-20
4	244	8.6	757	2	I45956 polymeric immunoglobulin	1.20e-19
5	223	7.9	773	1	QRBBG secretory component p	2.17e-16
6	140	4.9	224	2	CMRF-35 antigen - hum	1.75e-04
7	136	4.8	519	2	S14529 transition protein 2	5.73e-04
8	119	4.2	303	2	A44215 EUS1 protein - equine	7.29e-02
9	115	4.1	116	2	S04094 transition protein 2	2.16e-01
10	115	4.1	179	2	C36801 hypothetical protein	2.16e-01
11	117	4.1	567	2	C69611 ABC transporter requi	1.26e-01
12	115	4.1	1353	2	T00347 hypothetical protein	2.16e-01
13	113	4.0	136	2	PN0536 Ig heavy chain V regi	3.68e-01
14	114	4.0	146	1	G1HUH2 Ig heavy chain precu	2.82e-01
15	113	4.0	491	2	T12501 hypothetical protein	3.68e-01
16	113	4.0	863	1	S65954 replication licensing	3.68e-01
17	110	3.9	113	2	S57667 transition protein 2	8.11e-01
18	109	3.9	125	2	S37455 Ig mu chain - human	1.05e-00
19	109	3.9	502	2	S36494 E2 protein - human pa	1.05e-00
20	111	3.9	913	2	S20590 exo-alpha-sialidase (6.24e-01
21	109	3.9	2796	2	JC4743 fatty-acid synthase (1.05e-00
22	107	3.8	95	2	E47624 Ig heavy chain V-II r	1.76e-00
23	108	3.8	252	2	C43679 ORF3 protein - Autogr	1.36e+00

24	108	3.8	395	2	B72478 hypothetical protein	1.36e+00
25	108	3.8	699	2	T05225 extensin homolog F17I	1.36e+00
26	107	3.8	1333	2	T00249 ichl protein - inky c	1.76e+00
27	105	3.7	57	2	T03060 probable small basic	2.93e+00
28	104	3.7	114	2	T01262 immunoglobulin heavy	3.76e+00
29	104	3.7	129	2	S44114 Ig heavy chain V regi	3.76e+00
30	105	3.7	133	2	S34010 Ig heavy chain V regi	2.93e+00
31	105	3.7	134	2	S54906 Ig heavy chain V regi	2.93e+00
32	104	3.7	208	2	T16953 hypothetical protein	3.76e+00
33	104	3.7	256	2	A75013 hypothetical protein	3.76e+00
34	105	3.7	348	1	OZ2QBK circumsporozoite prot	2.93e+00
35	104	3.7	483	2	S36470 E2 protein - human pa	3.76e+00
36	105	3.7	508	2	T04605 hypothetical protein	2.93e+00
37	106	3.7	539	2	H71280 hypothetical protein	2.27e+00
38	105	3.7	575	2	S58647 vacuolar transport pr	2.93e+00
39	104	3.7	1404	2	E36788 hypothetical protein	3.76e+00
40	103	3.6	136	2	S58321 probable membrane pro	4.83e+00
41	102	3.6	354	2	A48931 transmembrane glycopr	6.19e+00
42	101	3.6	876	1	A57988 regulatory protein ar	6.19e+00
43	102	3.6	1203	2	S26650 DNA-binding protein 5	6.19e+00
44	102	3.6	1280	2	T00365 hypothetical protein	6.19e+00
45	103	3.6	1792	2	T08878 supervillin P205 - bo	4.83e+00

ALIGNMENTS

RESULT	1
ENTRY	QRUGS
TITLE	secretory component precursor - human
ALTERNATE_NAMES	poly-Ig receptor; polymeric immunoglobulin receptor
CONTAINS	free secretory component; transmembrane secretory component
ORGANISM	#formal_name Homo sapiens #common_name man
DATE	28-Aug-1985 #sequence_revision 23-Aug-1996 #text_change 16-Jul-1999
ACCESSIONS	A46537; A55284; I38115; A32263; S38978; S13453; A02112
REFERENCE	A46537
#authors	Krajci, P.; Kvale, D.; Tasken, K.; Brandtzaeg, P.
#journal	Eur. J. Immunol. (1992) 22:2309-2315
#title	Molecular cloning and exon-intron mapping of the gene encoding human transmembrane secretory component (the poly-Ig receptor).
#cross-references	MUID:92387236
#accession	A46537
#status	not compared with conceptual translation
#molecule_type	DNA
#residues	1-764 #label KRA
#cross-references	GB:S43449; NID:g255097; PIDN:AAB23176.1; PID:g255098
#experimental_source	leukocytes
#note	sequence extracted from NCBI backbone (NCBIP:113253)
REFERENCE	A55284
#authors	Krajci, P.; Grzeschik, K.H.; Geurts van Kessel, A.H.; Olaisen, B.; Brandtzaeg, P.
#journal	Hum. Genet. (1991) 87:642-648
#title	The human transmembrane secretory component (poly-Ig receptor): molecular cloning, restriction fragment length polymorphism and chromosomal sublocalization.
#cross-references	MUID:92039621
#accession	A55284
#molecule_type	mRNA
#residues	1-764 #label KR2
#cross-references	GB:S62403; NID:g238235; PIDN:AAB20203.1; PID:g238236
#experimental_source	colonic adenocarcinoma cell line
#note	sequence extracted from NCBI backbone (NCBIN:62403, NCBIP:62408)
REFERENCE	I38115
#authors	Piskurich, J.F.; France, J.A.; Tamer, C.M.; Willmer, C.A.; Kaezel, C.S.; Kaetzel, D.M.
#journal	Mol. Immunol. (1993) 30:413-421
#title	Interferon-gamma induces polymeric immunoglobulin receptor mRNA in human intestinal epithelial cells by a protein synthesis dependent mechanism.
#cross-references	MUID:93205018
#accession	I38115

```

##molecule_type mRNA
##residues 1-764 ##label RES
##cross-references EMBL:X73079; NID:9456345; PIDN:CAA51532.1;
PID:9456346
##note submitted to the EMBL/GenBank/DBJ databases by J.F.
Piskurich, February 1994

REFERENCE
#authors Krajci, P.; Solberg, R.; Sandberg, M.; Oyen, O.; Jahnsen, T.;
Brandtzaeg, P.
#journal Biochem Biophys. Res. Commun. (1989) 158:783-789
#title Molecular cloning of the human transmembrane secretory in
component (poly-Ig receptor) and its mRNA expression in
human tissues.
#cross-references MUID:89149795
#accession A32263
##molecule_type mRNA
##residues 72-764 ##label KR3
##cross-references GB:M24559; NID:9514365; PIDN:AAA36102.1; PID:9514366
S38978
#authors Fallgreen-Gebauer, E.; Gebauer, W.; Bastian, A.; Kratzin,
H.D.; Eifert, H.; Zimmermann, B.; Karas, M.; Hilschmann,
N.
#journal Biol. Chem. Hoppe-Seyler (1993) 374:1023-1028
#title The covalent linkage of secretory component to IgA. Structure
of sigA.
#cross-references MUID:94121784
#accession S38978
##molecule_type protein
##residues 478-488;517-526;543-545 ##label PAL
##note disulfide bonds for unbound and IgA-bound forms
S13453
#authors Eifert, H.; Quentin, E.; Wiederhold, M.; Hillemeir, S.;
Decker, J.; Weber, M.; Hilschmann, N.
#journal Biol. Chem. Hoppe-Seyler (1991) 372:119-128
#title Determination of the molecular structure of the human free
secretory component.
#cross-references MUID:91315750
#accession S13453
##molecule_type protein
##residues 19-135,'Q',137-157,'D',159-207,'DE',210-228,230-233,'N',
235-240,'Q',242-261,'Q',263-279,'N',281-391,'D',
393-499,'D',501-577 ##label E12
A02112
#authors Eifert, H.; Quentin, E.; Decker, J.; Hillemeir, S.;
Hufschmidt, M.; Klingmuller, D.; Weber, M.H.; Hilschmann,
N.
#journal Hoppe-Seyler's Z. Physiol. Chem. (1984) 365:1489-1495
#title The primary structure of the human free secretory component
and the arrangement of the disulfide bonds.
#cross-references MUID:85128981
#accession A02112
##molecule_type protein
##residues 19-157,'D',159-207,'DE',210-228,230-233,'N',235-240,'Q',
242-261,'Q',263-279,'N',281-391,'D',393-499,'D',
501-577 ##label E1F
#note paper in German with English abstract
COMMENT As a 100K transmembrane receptor for polymeric immunoglobulins,
secretory component is expressed on secretory epithelium
basolateral membranes. Upon binding IgA or IgM, it becomes
phosphorylated, forms interchain disulfide bonds, undergoes
proteolysis and transcytosis. Free secretory component may
protect polymeric immunoglobulins from proteolysis in mucosal
excretions.
GENETICS
#gene GDB:PIGR
##cross-references GDB:120290; OMIM:173880
#map_position 1q31-1q41
#introns 15/1; 130/1; 349/1; 460/1; 569/1; 629/2; 670/1; 714/1; 733/3
#note the first intron occurs before the initiator codon
monomeric as a transmembrane receptor or free in mucosal
secretions; heterodimer composed of one chain of
secretory component, one chain of immunoglobulin J chain,
and two IgA heterotetramers; hetero-22-mer composed of one
chain of secretory component, one chain of immunoglobulin J
chain, and five IgM heterotetramers
#superfamily secretory component; immunoglobulin homology
duplication; glycoprotein; immunoglobulin receptor;
phosphoprotein; transcytosis; transmembrane protein
FEATURE
1-18 #domain signal sequence #status predicted #label SIG\
19-764 #product transmembrane secretory component #status
#predicted #label MATW\
19-577 #product free secretory component #status experimental
#label MATF\
33-112 #domain immunoglobulin homology #label IM1\
145-222 #domain immunoglobulin homology #label IM2\
250-327 #domain immunoglobulin homology #label IM3\
364-443 #domain immunoglobulin homology #label IM4\
475-546 #domain immunoglobulin homology #label IM5\
639-661 #domain transmembrane #status predicted #label TMW\
682-764 #domain intracellular #status predicted #label INT\
40-110,56-64,
152-220,257-325,
271-279,371-441,
385-395,462-544,
496-503
83,90,135,186,421,
469,499
#disulfide_bonds #status experimental\
#binding_site carbohydrate (Asn) (covalent) #status
experimental\
#disulfide_bonds (in Ig-unbound form) #status
experimental\
#disulfide_bonds interchain (to IgA alpha-1 chain-192)
#status experimental\
#binding_site cysteine (Cys) (covalent) (in Ig-bound
form) #status experimental\
#cleavage_site Lys-Ala (unidentified proteinase) #status
experimental\
#binding_site phosphate (Ser) (covalent) #status
Predicted
SUMMARY #length 764 #molecular-weight 83313 #checksum 8299
Query Match 8.9%; Score 252; DB 1; Length 764;
Best Local Similarity 35.0%; Pred. No. 6.59e-21;
Matches 41; Conservative 30; Mismatches 40; Indels 6; Gaps 4;
Db 13 FPAISIKSPFGPEVNSVSGNSVITCYPTPTVNRHTRKYWCRO-GARGSGITLISSE 71
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 10 FLVPYSGALILPEVKYGEGLGSGVTKC---PLP-EMHVRILYLCREMGSGTCTGVVSTT 65
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 72 GYVSSKYAGRANLNFPEPNTFFVNTAQLSQDSDSGRYKCGLGINS-RLGSFDSVLEY 127
::: |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 66 NFIKAEYKRVTLKQYPRKNLFLVEVTQLTESDSGVYACGAGMNTDRGKTQKTVLNV 122
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
RESULT 2
ENTRY S48841 #type complete
TITLE secretory component precursor - bovine
ALTERNATE_NAMES poly-Ig receptor; polymeric immunoglobulin receptor
CONTAINS free secretory component; transmembrane secretory component
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
10-Sep-1999
ACCESSIONS S48841
REFERENCE S48841
#authors Vermeer, H.; Warner, G.W.; de Boer, H.A.; Verbeet, M.P.H.
#submission submitted to the EMBL Data Library, September 1994
#description The cloning, tissue specific expression and interspecies
sequence comparison of the bovine polymeric immunoglobulin
receptor.
#accession S48841
#status preliminary
#molecule_type mRNA
#residues 1-757 #label VER
#cross-references EMBL:X81371; NID:9563340; PIDN:CAA57136.1;
PID:9563341
CLASSIFICATION #superfamily secretory component; immunoglobulin homology

```

KEYWORDS duplication; glycoprotein; immunoglobulin receptor; phosphoprotein; transcytosis; transmembrane protein

FEATURE 1-18 #domain signal sequence #status predicted #label SIG\ 19-757 #product transmembrane secretory component #status predicted #label MATM\ 19-572 #product free secretory component #status predicted #label MATM\ 33-112 #domain immunoglobulin homology #label IM1\ 145-222 #domain immunoglobulin homology #label IM2\ 250-328 #domain immunoglobulin homology #label IM3\ 365-444 #domain immunoglobulin homology #label IM4\ 476-547 #domain immunoglobulin homology #label IM5\ 631-653 #domain transmembrane #status predicted #label TM\ 654-757 #domain intracellular #status predicted #label INT\ 40-110,56-64, 152-220,166-173, 257-324,271-279, 370-440,384-394, 481-543,495-502 83,420,468

665 #disulfide_bonds #status predicted\ #binding_site carbohydrate (Asn) (covalent) #status predicted\ #binding_site phosphate (Ser) (covalent) #status predicted\ #length 757 #molecular-weight 82450 #checksum 7998

SUMMARY Query Match 8.7%; Score 245; DB 1; Length 757; Best Local Similarity 36.7%; Pred. No. 8.34e-20; Matches 40; Conservative 25; Mismatches 37; Indels 7; Gaps 5;

Db 25 PE-EVSSVEGRSVTKYVPPTSVNRHTRKYWCRO-GAOGRCTTLISSEGYSDDYVGRA 82
 QY 21 PEVKVEGLGSGVTIKC---PLP-EMHVRILCREMAGSGTCGTVTTFIKAEYKGRV 76

Db 83 NLTFNPESGTFVVDISHTHKDSGRYKGLGISS-RGLNFDVSLEVSQD 130
 QY 77 TLKOYPRKNLFLVEVTQLTSDSGYACGAGMNTDRGKTQKVTNLNVHSE 125

RESULT 3 QRTGS #type complete ENTRY secretory component precursor - rat ALTERNATE_NAMES poly-ig receptor; polymeric immunoglobulin receptor CONTAINS free secretory component; transmembrane secretory component ORGANISM #formal_name Rattus norvegicus #common_name Norway rat DATE 07-Sep-1990 #sequence_revision 23-Aug-1996 #text_change 16-Jul-1999

ACCESSIONS S05407; S54731 REFERENCE S05407 Banting, G.; Brake, B.; Braghetta, P.; Luzio, J.P.; Stanley, K.K. FEBS Lett. (1989) 254:177-183 Intracellular targetting signals of polymeric immunoglobulin receptors are highly conserved between species. #cross-references EMBL:89378226 #accession S05407 #molecule_type mRNA #residues 1-769 #label BAN #cross-references EMBL:X15741; NID:g56464; PIDN:CRA33758.1; PID:g56465 monomeric as a transmembrane receptor or free in mucosal secretions; heterodimer composed of one chain of secretory component, one chain of immunoglobulin J chain, and two IgA heterotetramers; hetero-22-mer composed of one chain of secretory component, one chain of immunoglobulin J chain, and five IgM heterotetramers

CLASSIFICATION #superfamily secretory component; immunoglobulin homology duplication; glycoprotein; immunoglobulin receptor; phosphoprotein; transcytosis; transmembrane protein

KEYWORDS #domain signal sequence #status predicted #label SIG\ 1-18 #product transmembrane secretory component #status predicted #label MATM\ 19-769 #product free secretory component #status predicted

33-112 #label MATF\ #domain immunoglobulin homology #label IM1\ 145-222 #domain immunoglobulin homology #label IM2\ 250-326 #domain immunoglobulin homology #label IM3\ 363-442 #domain immunoglobulin homology #label IM4\ 477-548 #domain immunoglobulin homology #label IM5\ 644-666 #domain transmembrane #status predicted #label TM\ 667-769 #domain intracellular #status predicted #label INT\ 40-110,56-64, 152-220,370-440, 384-394,484-546, 498-505 90,135,471

488-522 #disulfide_bonds #status predicted\ #binding_site carbohydrate (Asn) (covalent) #status predicted\ #disulfide_bonds interchain (to IgA alpha-1 chain-192) #status predicted\ #binding_site cysteine (Cys) (covalent) (in Ig-bound form) #status predicted\ #binding_site phosphate (Ser) (covalent) #status predicted\ #length 769 #molecular-weight 84798 #checksum 7578

SUMMARY Query Match 8.7%; Score 246; DB 1; Length 769; Best Local Similarity 35.0%; Pred. No. 5.81e-20; Matches 41; Conservative 30; Mismatches 40; Indels 6; Gaps 5;

Db 13 FSGYSTOSPIFGPODVSSIEGNSVITCYPTDTSVNRHTRKYWCRO-GANGYCALIISN 71
 QY 10 FLVPVSGALRLPELVKEVGEGLGSGVTIKC--P-LP-EMHVRILCREMAGSGTCGTIVSTT 65

Db 72 GYLSKEVSGRASLNFNFENSTFVNIAHLTQEDTGSYKGLG-TTNRGLFFDVSLEV 127
 QY 66 NFIRAEYKGRVTLKQYPRKNLFLVEVTQLTSDSGYACGAGMNTDRGKTQKVTNLNV 122

RESULT 4 I45956 #type complete ENTRY polymeric immunoglobulin receptor - bovine TITLE #formal_name Bos primigenius taurus #common_name cattle ORGANISM 19-Dec-1997 #sequence_revision 19-Dec-1997 #text_change 23-Jul-1999

ACCESSIONS I45956 REFERENCE I45956 Kulsath, M.A.; Krajci, P.; Myklebost, O.; Rogne, S. DNA Cell Biol. (1995) 14:251-256 Cloning and characterization of two forms of bovine polymeric immunoglobulin receptor cDNA. #cross-references MUID:95186063 #accession I45956 #status preliminary; translated from GB/EMBL/DBJ #molecule_type mRNA #residues 1-757 #label KUL #cross-references GB:I04797; NID:g388279; PIDN:AAC41620.1; PID:g388280 CLASSIFICATION #superfamily secretory component; immunoglobulin homology FEATURE 145-222 #domain immunoglobulin homology #label IM\ 145-222 #length 757 #molecular-weight 82434 #checksum 8761

SUMMARY Query Match 8.6%; Score 244; DB 2; Length 757; Best Local Similarity 37.0%; Pred. No. 1.20e-19; Matches 37; Conservative 24; Mismatches 33; Indels 6; Gaps 4;

Db 33 GRSVSIKCYKYPPTSVNRHTRKYWCRO-GAOGRCTTLISSEGYSDDYVGRANLTFPESG 91
 QY 30 GGSVTIKC---PLP-EMHVRILCREMAGSGTCGTIVSTTFNFIKAEYKGRVTLKQIPRKN 85

Db 92 TFVVDISHLTHKDSGRYKGLGISS-RGLNFDVSLEVSQD 130
 QY 86 LFLVEVTQLTSDSGYACGAGMNTDRGKTQKVTNLNVHSE 125

RESULT 5

```

ENTRY
  ORRG      #type complete
  TITLE     Secretory component precursor - rabbit
  ALTERNATE_NAMES
  CONTAINS  free secretory component; transmembrane secretory component
  ORGANISM  #formal_name Oryctolagus cuniculus #common_name domestic
            rabbit
  DATE      15-Nov-1984 #sequence_revision 15-Nov-1984 #text_change
            16-Jul-1999
  ACCESSIONS
  REFERENCE
  #authors  A02111; A28077
  #journal  Mostov, K.E.; Friedlander, M.; Blobel, G.
  #title    Nature (1984) 308:37-43
  #cross-references M01D:8412246
  #molecule_type mRNA
  #residues 1-773 #label MOS
  #cross-references GB:X00412; GB:X01291; NID:g1595; PIDN:CAA25118.1;
            PID:g1596
  #note     the authors translated the codon ACC for residue 54 as
            Asn
  REFERENCE
  #authors  A28077
  #journal  Frutiger, S.; Hughes, G.J.; Hanly, W.C.; Jaton, J.C.
  #title    J. Biol. Chem. (1988) 263:8420-8125
  #cross-references M01D:88228032
  #molecule_type protein
  #residues 87-114:410-424 #label FRU
  COMMENT  This receptor binds polymeric IGA and IGM at the basolateral
            surface of epithelial cells. The complex is then transported
            across the cell to be secreted at the apical surface. During this
            process, cleavage occurs to separate the extracellular portion,
            also known as the secretory component, from the transmembrane
            segment.
  COMMENT  The five domains exhibit homology with immunoglobulin V regions.
            The similarity is strongest between the fourth domain and kappa
            chain V regions.
  COMMENT  Alternative splicing in the extracellular domain leads to high or
            low molecular weight forms of secretory component.
  CLASSIFICATION #superfamily secretory component; immunoglobulin homology
  KEYWORDS  alternative splicing; duplication; glycoprotein;
            immunoglobulin receptor; polymorphism; transcytosis;
            transmembrane protein
  FEATURE
  1-18      #domain signal sequence #status predicted #label SIG\
  19-773    #product transmembrane secretory component #status
            predicted #label MATM\
  19-575    #product free secretory component #status predicted
            #label MATF\
  30-647    #domain extracellular #status predicted #label EXT\
  39-117    #domain immunoglobulin homology #label IM1\
  148-227   #domain immunoglobulin homology #label IM2\
  253-326   #domain immunoglobulin homology #label IM3\
  362-440   #domain immunoglobulin homology #label IM4\
  471-540   #domain immunoglobulin homology #label IM5\
  648-670   #domain transmembrane #status predicted #label TM\
  671-773   #domain intracellular #status predicted #label INT\
  46-115,155-225,
  260-324,369-438,
  478-538   #disulfide_bonds #status predicted\
  108       #binding_site carbohydrate (Asn) (covalent) (partial)
            #status experimental\
  418       #binding_site carbohydrate (Asn) (covalent) #status
            experimental
  SUMMARY   #length 773 #molecular-weight 83886 #checksum 7723
            7.9%; Score 223; DB 1; Length 773;
  Query Match
  Best Local Similarity 45.5%; Pred. No. 2.17e-16;
  Matches 40; Conservative 11; Mismatches 30; Indels 7; Gaps 5;

```

```

Db 39 GDSVSYICYPPTSVTRHSRKFRCRE-ESGRHCVTLAST-GYTSEYSGRGLTDFPKG 96
QY 30 GGSVTIKC--PLPEM--HVRIYLCREMAGSGCGTGVSTTTFIKAIEYGRVTLKQYPRKN 85
Db 97 EFVTVDTQLTQNDSGSYRCGVGVN-GRG 123
QY 86 LFLVEVTQLTESDSGVYACGAGMNTDRG 113

RESULT 6
ENTRY   I37243      #type complete
TITLE   CRP-35 antigen - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE     12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
            28-Feb-1997
ACCESSIONS
REFERENCE I37243
#authors  Jackson, D.G.; Hart, D.N.; Starling, G.; Bell, J.I.
#journal  Eur. J. Immunol. (1992) 22:1157-1163
#title    Molecular cloning of a novel member of the immunoglobulin
            gene superfamily homologous to the polymeric immunoglobulin
            receptor.
#cross-references M01D:92249405
#accession I37243
#status    preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-224 #label RES
#cross-references EMBL:X66171; NID:g396169; PID:g396170
GENETICS
#gene     CMRF35
SUMMARY   #length 224 #molecular-weight 24830 #checksum 8067
            4.9%; Score 140; DB 2; Length 224;
            Best Local Similarity 29.7%; Pred. No. 1.75e-04;
            Matches 30; Conservative 23; Mismatches 44; Indels 4; Gaps 3;
Db 13 ALLLLVPGVPLSHPMVAGVGGISVOCHEKEHRTLNKEWGRP-PQILRCDKIVET 71
QY 7 PLFLPVGALRILPEVVEGEGSVTIKPLPEMH--VRIYLCREMAGSGTGVST 64
Db 72 KGSA-GRKRVSRVSRDSPANLSTFTVLENTEDAGTYWCG 111
QY 65 TNFIKAYKRGVTLKQYPRKNLFLVEVTQLTESDSGVYACG 105

RESULT 7
ENTRY   S14529      #type complete
TITLE   transition protein 2 - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE     20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
            16-Jul-1999
ACCESSIONS
REFERENCE S14529; JQ0801; A28447
#authors  Nelki, S.; Greenhall, C.; Shanahan, F.; Dudley, K.
#journal  submitted to the EMBL Data Library, November 1989
#description Identification and DNA sequence of a family of related cDNA's
            one of which is haploid specific and has complete homology.
#accession S14529
#status    preliminary
#molecule_type mRNA
#residues 1-519 #label NEL
#cross-references EMBL:X17068; NID:g54873; PIDN:CAA34914.1; PID:g54874
REFERENCE JQ0801
#authors  Kleene, K.C.; Gerstel, J.; Shih, D.
#journal  Gene (1990) 95:301-302
#title    Nucleotide sequence of the gene encoding mouse transition
            protein 2.
#cross-references M01D:91065546
#accession JQ0801
#molecule_type DNA
#residues 403-512, 'T', 514-519 #label KLE
#cross-references GB:M60254; GB:M36193; NID:g202107; PIDN:AAA0469.1;

```


2

#cross-references	MUID:98044033
#accession	C69611
#status	preliminary; nucleic acid sequence not shown;
	translation not shown
#molecule_type	DNA
#residues	1-567 #label KUN
#cross-references	GB:AL009126; NID:g2636240; PID:e1186373;
#experimental_source	strain 168
GENETICS	
#gene	cycD
CLASSIFICATION	#superfamily ATP-binding cassette homology
KEYWORDS	ATP; p-loop
FEATURE	
343..537	#domain ATP-binding cassette homology #label ABC\
360..367	#region nucleotide-binding motif A (p-loop)
SUMMARY	#length 567 #molecular-weight 62805 #checksum 9210
Query Match	4.1%; Score 117; DB 2; Length 567;
Best Local Similarity	31.3%; Pred.No. 1.26e+01;
Matches	20; Conservative 21; Mismatches 20; Indels 3; Gaps 3;
Db	158 ILVAAMPILIFILLGLVQAQRKADRWKSYQZRSNHFVDLSRLGET-LRFLGLSKSHSK 216
	::: :: :: :: :: :: :: :: :: :: :: ::
QY	254 ILIPITLGLF-LALLGLVKRAVER-RKALSRRARLRVMRALESSQRGSPRPQRSQ 311
	::: :: :: :: :: :: :: :: :: :: :: ::
Db	217 NIFY 220
	::
QY	312 NNYY 315
RESULT	12
ENTRY	T00347
TITLE	hypothetical protein DKFzP566G1246.1, version I - human
ALTERNATE_NAMES	(fragment)
ORGANISM	hypothetical protein KIAA0592
DATE	#formal_name Homo sapiens #common_name man
	01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change
ACCESSION	T00347
REFERENCE	Z14086
#authors	Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
#journal	DNA RES. (1998) 5:31-39
#title	Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.
#cross-references	MUID:98290545
#accession	T00347
#status	translated from GB/EMBL/DBJ
#molecule_type	mRNA
#residues	1-1353 #label NAG
#cross-references	EMBL:AB011164; NID:d1185434; PID:d1026448
#experimental_source	tissue brain; clone HU2807
GENETICS	
#note	KIAA0592
SUMMARY	#length 1353 #checksum 6126
Query Match	4.1%; Score 115; DB 2; Length 1353;
Best Local Similarity	27.4%; Pred.No. 2.16e+01;
Matches	43; Conservative 40; Mismatches 66; Indels 8; Caps 7;
Db	976 NPAAALP-TAASOISVRKPVLPPELAAPSEHRSRGLESVPVLPGSGEAGVSD-L-PAQ 1032
	: :: :: :: :: :: :: :: :: :: :: :: ::
QY	200 KPRTFLPTTAKSIALGLKLPOTPSYNHHTRLRHQRALDYQSOGREGGFPHLIPTI 259
	: :: :: :: :: :: :: :: :: :: :: :: ::
Db	1033 ADTLHSANKSVKWRG-KRRPQTQRAARRLAAQESSAEADMVSITPRGTQAOWADGISPNQ 1090
	: :: :: :: :: :: :: :: :: :: :: :: ::
QY	260 LGLFLALLGLLVKRAVERRKALSRRARLRVMREA-LESSQRGSPRPQRSQNIIYSAC 318
	: :: :: :: :: :: :: :: :: :: :: :: ::
Db	1091 HRPOLRAASGDSTEEALAAAAAWEVGPPGVDTSP 1127
	: :: :: :: :: :: :: :: :: :: :: :: ::

[illegible]

[W][O][R][L][D]
[D][I][S][T][R][I][B][U][T][I][O][N]
[R][I][G][H][T][S]
[B][Y]
[O][X][F][O][R][D]
[M][O][L][E][C][U][L][A][R]
[L][I][M][I][T][E][D]

(TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Aug 24 10:24:36 2000; Maspar time 12.26 Seconds
Tabular output not generated. 753.292 Million cell updates/sec

Title: >US-09-135-238B-2
Description: (1-390) from US09135238B.pep
Perfect Score: 2830
Sequence: 1 MDRWLWPLFLVFLVSGALRL.....HQPAMMEDSDSDYINVPA 390

Scoring table: PAM 150
Gap 11

Searched: 188963 segs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 36.230; Variance 187.638; scale 0.193

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	2815	99.5	390	1 Y05001	Human PIGRL-1 protein	5.91e-218
2	787	27.8	107	1 Y05002	Human PIGRL-1 protein	3.36e-51
3	254	9.0	771	1 W03180	Mouse poly-immunoglobulin	2.06e-09
4	252	8.9	607	1 W95601	Human secretory immunoglobulin	2.90e-09
5	247	8.7	746	1 W03178	Human poly-immunoglobulin	6.77e-09
6	245	8.7	757	1 W03179	Bovine poly-immunoglobulin	9.49e-09
7	246	8.7	769	1 W03181	Rat poly-immunoglobulin	8.01e-09
8	223	8.0	532	1 W50033	Human immunoglobulin related	2.32e-07
9	223	7.9	584	1 R14670	Truncated poly Ig-rece	3.83e-07
10	223	7.9	773	1 W03177	Rabbit poly-immunoglobulin	3.83e-06
11	209	7.4	584	1 R14671	Truncated poly Ig-rece	3.92e-06
12	167	5.9	205	1 Y05069	Human PIGR-2 protein s	3.54e-03
13	129	4.6	201	1 W99070	Human PIGR-1	1.23e+00
14	130	4.6	211	1 W86306	Kidney injury associat	1.06e+00
15	125	4.4	195	1 W99077	Immunoglobulin like pr	2.23e+00
16	123	4.3	195	1 W99071	Partial human PIGR-1	2.99e+00
17	122	4.3	303	1 R84103	Equine herpesvirus (EH	3.46e+00
18	122	4.3	303	1 R52667	Equine herpesvirus US2	3.46e+00
19	118	4.2	319	1 W14146	Human A33 antigen	6.19e+00
20	112	4.0	57	1 W99072	Immunoglobulin like pr	1.47e+01
21	112	4.0	60	1 W99073	Immunoglobulin like pr	1.47e+01
22	113	4.0	83	1 Y04776	Mycobacterium species	1.27e+01
23	113	4.0	104	1 Y04773	Mycobacterium species	1.27e+01

24	113	4.0	134	1 R04385	Colon Cancer monoclonal	1.27e+01
25	113	4.0	134	1 R38314	Sequence of murine ant	1.27e+01
26	113	4.0	139	1 W59615	Anti-RSV F protein Hul	1.27e+01
27	109	3.9	114	1 R48617	Sequence of the monoc	2.25e+01
28	110	3.9	138	1 W59617	Anti-RSV F protein Hul	1.95e+01
29	111	3.9	476	1 W01818	Primate anti-human	1.69e+01
30	111	3.9	476	1 W63761	Macaque primatized 7C1	1.69e+01
31	108	3.8	120	1 R50216	HSV glycoprotein F bin	2.59e+01
32	108	3.8	138	1 W59616	Anti-RSV F protein Hul	2.59e+01
33	107	3.8	138	1 R32246	BR55-2 murine IgG3 hea	2.98e+01
34	107	3.8	138	1 R32242	Chimeric MAB heavy cha	2.98e+01
35	107	3.8	138	1 W59618	Anti-RSV F protein Hul	2.98e+01
36	107	3.8	139	1 R31588	BR55-2 heavy chain var	2.98e+01
37	108	3.8	309	1 R83322	Single chain Apo-2 ant	2.59e+01
38	108	3.8	377	1 W53751	Seven transmembrane re	2.59e+01
39	108	3.8	377	1 W48732	Human R2 seven transme	2.59e+01
40	107	3.8	659	1 W11727	Delta-1 polypeptide	2.98e+01
41	106	3.7	97	1 R50100	Sequence encoded by hu	3.42e+01
42	106	3.7	118	1 R68739	MAB B3 heavy chain	3.42e+01
43	106	3.7	247	1 R32091	H and L chain FV regio	3.42e+01
44	106	3.7	249	1 R95213	Heavy and light chain	3.42e+01
45	106	3.7	249	1 R32090	H and L chain FV regio	3.42e+01

ALIGNMENTS

RESULT 1
ID Y05001 standard; Protein; 390 AA.
AC Y05001;
DT 16-JUN-1999 (first entry)
DE Human PIGRL-1 protein sequence.
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;
KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;
KW diagnosis; therapy.
OS Homo sapiens.
PN EP-905238-A2.
PD 31-MAR-1999.
PF 14-AUG-1998; 306487.
PR 30-OCT-1997; US-961564.
PR 25-AUG-1997; US-056935.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Sweet RW, Truneh A, Wu S;
DR WPI; 99-192666/17.
DR N-PSDB; X28178.
PT New polypeptides encoding PIGRL-1 useful for treating diseases such
PT as X-linked Severe Combined Immunodeficiency
PS Claim 11; Page 7; 26pp; English.
CC This sequence is the human PIGRL-1 protein of the invention.
CC Autoimmune diseases involving altered expression or activity of PIGRL-1
CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined
CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be
CC diagnosed or susceptibility to them predicted by: (1) determining whether
CC there is a mutation in the genomic copy of the gene encoding PIGRL-1;
CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.
CC Patients deficient in PIGRL-1 can be treated by administering either the
CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.
CC Patients with excessive expression or activity of PIGRL-1 can be treated
CC by administering an antagonist of PIGRL-1, an antisense nucleic acid
CC molecule which inhibits the expression of PIGRL-1 or administering
CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can
CC be used to identify its agonists by contacting a cell expressing PIGRL-1
CC with a candidate compound in the presence of a signal system and noting
CC the candidate as an agonist if a signal is produced. The same method can
CC be used to identify antagonists of PIGRL-1 but the presence of an
CC antagonist is indicated by a decrease in production of the signal.
CC Antibodies against PIGRL-1 may be used to isolate or identify clones
CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to
CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.
CC This information may then be correlated with the incidence of autoimmune
CC disease in those patients to identify whether the mutation causes the
CC disease.
SQ Sequence 390 AA;

Query Match 99.5%; Score 2815; DB 1; Length 390;
 Best Local Similarity 99.7%; Pred. No. 5.91e-218;
 Matches 389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 MDFWLPLYFLPVSGALRILPEVKVEGELGSGVTIKCPLPMHVRVYLCREMAGSGTCGT 60
 |||||||
 QY 1 MDRWLPLYFLPVSGALRILPEVKVEGELGSGVTIKCPLPMHVRVYLCREMAGSGTCGT 60
 |||||||

Db 61 VYSTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTESDSGVYACG 120
 |||||||
 QY 61 VYSTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTESDSGVYACG 120
 |||||||

Db 121 NVHSEYEPSWEEQMPETPKWFLPYLFQMPAYASSSKFVTRVTTTQAQKGKVPVPHSSP 180
 |||||||
 QY 121 NVHSEYEPSWEEQMPETPKWFLPYLFQMPAYASSSKFVTRVTTTQAQKGKVPVPHSSP 180
 |||||||

Db 181 TQTLTHPRVSRASSVAGDKPRTFLPSTTASKISALEGLLKPTQPSYNHHTLHRQALD 240
 |||||||
 QY 181 TQTLTHPRVSRASSVAGDKPRTFLPSTTASKISALEGLLKPTQPSYNHHTLHRQALD 240
 |||||||

Db 241 YGSGREGGPHILPTILGLFLALLGLVVKRAVERKALSRARRLAVMRALLESSQ 300
 |||||||
 QY 241 YGSGREGGPHILPTILGLFLALLGLVVKRAVERKALSRARRLAVMRALLESSQ 300
 |||||||

Db 301 RPRGSPRPSQNNIYSACPRARCAAGTGEAPVPGGAPLPAPLQVSSPWLHAPSL 360
 |||||||
 QY 301 RPRGSPRPSQNNIYSACPRARCAAGTGEAPVPGGAPLPAPLQVSSPWLHAPSL 360
 |||||||

Db 361 KTCSEYVSLYHOPAAAMMEDSDSDYINVPA 390
 |||||||
 QY 361 KTCSEYVSLYHOPAAAMMEDSDSDYINVPA 390
 |||||||

RESULT 2

ID Y05002 standard; peptide; 107 AA.
 AC Y05002;
 DE 16-JUN-1999 (first entry)
 DT Human PIGRL-1 protein sequence fragment.
 KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;
 KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;
 KW diagnosis; therapy.
 OS Homo sapiens.
 PN EP-905238-A2.
 PD 31-MAR-1999.
 PF 14-AUG-1998; 306487.
 PR 30-OCT-1997; US-961564.
 PR 25-AUG-1997; US-056935.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Sweet RW, Truneh A, Wu S;
 DR WPI: 99-192666/17.
 DR N-PSDB; X28179.
 FT New polypeptides encoding PIGRL-1 useful for treating diseases such
 PT as X-linked Severe Combined Immunodeficiency
 PS Disclosure; Page 9; 26pp; English.
 CC This sequence is a fragment of the human PIGRL-1 of the invention.
 CC Autoimmune diseases involving altered expression or activity of PIGRL-1
 CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined
 CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be
 CC diagnosed or susceptibility to them predicted by: (1) determining whether
 CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or
 CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.
 CC Patients deficient in PIGRL-1 in a sample derived from the patient.
 CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.
 CC Patients with excessive expression or activity of PIGRL-1 to the patient.
 CC by administering an antagonist of PIGRL-1, an antisense nucleic acid
 CC molecule which inhibits the expression of PIGRL-1 or administering
 CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can
 CC be used to identify its agonists by contacting a cell expressing PIGRL-1
 CC with a candidate compound in the presence of a signal system and noting
 CC the candidate as an agonist if a signal is produced. The same method can
 CC be used to identify antagonists of PIGRL-1 but the presence of an
 CC antagonist is indicated by a decrease in production of the signal.
 CC Antibodies against PIGRL-1 may be used to isolate or identify clones

CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to
 CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.
 CC This information may then be correlated with the incidence of autoimmune
 CC disease in those patients to identify whether the mutation causes the
 CC disease.
 CC Sequence 107 AA;

Query Match 27.8%; Score 787; DB 1; Length 107;
 Best Local Similarity 99.0%; Pred. No. 3.36e-51;
 Matches 104; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 MDFWLPLYFLPVSGALRILPEVKVEGELGSGVTIKCPLPMHVRVYLCREMAGSGTCGT 60
 |||||||
 QY 1 MDRWLPLYFLPVSGALRILPEVKVEGELGSGVTIKCPLPMHVRVYLCREMAGSGTCGT 60
 |||||||

Db 61 VYSTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTESDSGVYACG 105
 |||||||
 QY 61 VYSTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTESDSGVYACG 105
 |||||||

RESULT 3

ID W03180 standard; Protein; 771 AA.
 AC W03180;
 DT 24-FEB-1997 (first entry)
 DE Mouse poly-immunoglobulin receptor.
 KW Mouse; immunoglobulin; receptor; protection protein; mutans;
 KW heavy chain; antigen binding domain; protection; pathogen;
 KW mucosal; environment; gastrointestinal; passive; immunisation;
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
 KW poly; sorbinus; murine.
 OS Mus musculus.
 FH Key
 FT region 13..45
 FT /note= "putative immunoglobulin binding residues
 FT of domain I"
 FT domain 1..120
 FT /label= domain_I
 FT domain 110..230
 FT /label= domain_II
 FT domain 210..340
 FT /label= domain_III
 FT domain 320..450
 FT /label= domain_IV
 FT domain 440..550
 FT /label= domain_V
 FT region 550..606
 FT /note= "external portions of domain VI"
 FT region 550..627
 FT /note= "external portions of domain VI"
 FT region 625..660
 FT /label= transmembrane_segment
 FT region 650..771
 FT /label= intracellular_portion
 DN W0921012-A1.
 PD 11-JUL-1996.
 PF 27-DEC-1995; U16889.
 PR 30-DEC-1994; US-367395.
 PR 04-MAY-1995; US-434000.
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.
 PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
 PA (PLAN-) PLANET BIOTECHNOLOGY INC.
 PI Hiatt AC, Lehner T, Ma JKC;
 DR WPI: 96-333987/33.
 DR N-PSDB; T31290.
 FT Immunoglobulin and protection protein complex and its prodn. in
 PT plants - useful for passive immunisation against mucosal antigens,
 PT esp. against S. mutans and S. sorbinus to prevent dental caries
 PS Disclosure; Pages 117-121; 152pp; English.
 CC The present sequence is the mouse poly-immunoglobulin (Ig)
 CC receptor, a portion of which correspo. to residues 1-627, pref.
 CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
 CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
 CC The Ig of the invention comprises a PP as above in association with

CC an Ig derived heavy chain, having at least a portion of an antigen
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
 CC gastrointestinal, environments, therefore enhancing its
 CC effectiveness in passively immunising animals against mucosal
 CC pathogens. The Ag binding domain is specifically derived from the
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
 CC S. sorbinus serotypes d and g.
 SQ Sequence 771 AA;

Query Match 9.0%; Score 254; DB 1; Length 771;
 Best Local Similarity 36.8%; Pred. No. 2.06e-09;
 Matches 43; Conservative 26; Mismatches 42; Indels 6; Gaps 5;

Db 13 FSGVSTKPIGPQVEISIEGDSVITCYPPPTSVNRHTRKYWCRO-GASGMCTTLLSN 71
 QY 10 FLFVSGALRIILPEVKVGEGLGGVTIKC--P-LP-EHVRVYILCREWAGSGTCGVVST 65
 Db 72 GYLKSEYSGRANLNFNPENFTVINISQLQDGTGSKYKGLG-TSNRGLSFDVSLV 127
 QY 66 NFIRAKEYGRVTLKQYPRKNLFLVEVTLQTESDGVYACGAGMNTDRGKTQKVTLLNV 122

RESULT 4

ID W95601 standard; Protein; 607 AA.
 AC W95601:
 DE 08-JUN-1999 (first entry)
 DT Human secretory immunoglobulin A component.
 KW Immunoglobulin A; secretory; component; IgA; human; treatment;
 KW prevention; infection; HIV; AIDS; cold; flu; virus;
 KW human immunodeficiency virus; respiratory syncytial virus.
 OS Homo sapiens.
 PN W09857993-A1.
 PD 23-DEC-1998.
 PF 10-JUN-1998; U11975.
 PR 19-JUN-1997; US-050969.
 PA (REGC) UNIV CALIFORNIA.
 PI Chintalacharuvu KR, Morrison SL;
 DI WPI: 99-080950/07.
 DR N-PSDB: X07407.
 PT Producing secretory immunoglobulin in single cells - useful to
 PT produce commercial quantities of secretory immunoglobulin to prevent
 PT or treat infections
 PS Disclosure; Pages 22-24; 39pp; English.
 CC The sequence is that of the secretory component of human secretory
 CC immunoglobulin A (sIgA). It can be used as part of a method for
 CC the production of sIg molecules. This method is useful for
 CC producing commercial quantities of sIg (especially sIgA) to treat
 CC or prevent infections. In particular, sIgA produced by the method
 CC can be used to prevent or treat infections in mammals, birds or
 CC fish; especially systemic infections or infections at a mucosal
 CC surface. It is especially useful to prevent or treat infection
 CC with human immunodeficiency virus (HIV), respiratory syncytial
 CC virus, flu virus or cold virus. The method allows production of
 CC commercial quantities of sIg molecules for therapeutic use, not
 CC previously possible; production using non-plant cells and a
 CC single cell type is more efficient than a previous multi-step
 CC process of fusing recombinant plant cells, and avoids alterations
 CC of the sIg by plant cells. sIgA molecules are more stable
 CC and resistant to proteolysis than previously used IgA molecules,
 CC and can be administered to prevent as well as to treat infections,
 CC unlike e.g. IgG and IgM molecules.
 SQ Sequence 607 AA;

Query Match 8.9%; Score 252; DB 1; Length 607;
 Best Local Similarity 35.0%; Pred. No. 2.90e-09;
 Matches 41; Conservative 30; Mismatches 40; Indels 6; Gaps 4;

Db 13 FPAITKSPIGPEVNSVEGNSVITCYPPPTSVNRHTRKYWCRO-GARGGCITLISSE 71
 QY 10 FLFVSGALRIILPEVKVGEGLGGVTIKC--P-LP-EHVRVYILCREWAGSGTCGVVST 65
 Db 72 GYVSSKYAGRANLTFNPENFTVINISQLQDGTGSKYKGLGINS-RGLSFDVSLV 127

QY 66 NFIRAKEYGRVTLKQYPRKNLFLVEVTLQTESDGVYACGAGMNTDRGKTQKVTLLNV 122

RESULT 5

ID W03178 standard; Protein; 746 AA.
 AC W03178:
 DT 24-FEB-1997 (first entry)
 DE Human poly-immunoglobulin receptor.
 KW Human; immunoglobulin; receptor; protection protein; mutans;
 KW heavy chain; antigen binding domain; protection; pathogen;
 KW mucosal; environment; gastrointestinal; passive; immunisation;
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
 KW poly; sorbinus.
 OS Homo sapiens.
 PN Key
 FT region
 FT 13..45 Location/Qualifiers
 FT /note="putative immunoglobulin binding residues
 FT of domain I"
 FT 1..120
 FT /label="domain_I
 FT 110..230
 FT /label="domain_II
 FT 210..340
 FT /label="domain_III
 FT 320..450
 FT /label="domain_IV
 FT 440..550
 FT /label="domain_V
 FT 550..606
 FT /note="external portions of domain VI"
 FT 550..627
 FT /note="external portions of domain VI"
 FT 625..660
 FT /label="transmembrane_segment
 FT 650..746
 FT /label="intracellular_portion
 PN W09621012-A1.
 PD 11-JUL-1996.
 PF 27-DEC-1995; U16889.
 PR 30-DEC-1994; US-367395.
 PR 04-MAY-1995; US-434000.
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.
 PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
 PA (PLAN-) PLANET BIOTECHNOLOGY INC.
 PI Hiatt AC, Lehner T, Ma JKC;
 DI WPI: 96-333987/33.
 DR N-PSDB: T31288.
 PT Immunoglobulin and protection protein complex and its prodn. in
 PT plants - useful for passive immunisation against mucosal antigens,
 PT esp. against S. mutans and S. sorbinus to prevent dental caries
 PS Disclosure; Pages 105-108; 152pp; English.
 CC The present sequence is the human poly-immunoglobulin (Ig)
 CC receptor, a portion of which corresp. to residues 1-627, pref.
 CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
 CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
 CC The Ig of the invention comprises a PP as above in association with
 CC an Ig derived heavy chain, having at least a portion of an antigen
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
 CC gastrointestinal, environments, therefore enhancing its
 CC effectiveness in passively immunising animals against mucosal
 CC pathogens. The Ag binding domain is specifically derived from the
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
 CC S. sorbinus serotypes d and g.
 SQ Sequence 746 AA;

Query Match 8.7%; Score 247; DB 1; Length 746;
 Best Local Similarity 37.7%; Pred. No. 6.77e-09;
 Matches 40; Conservative 27; Mismatches 32; Indels 7; Gaps 5;
 Db 7 PE-EVNSVEGNSVITCYPPPTSVNRHTRKYWCRO-GARGGCITLISSEGYVSSKYAGRA 64

QY 21 PEVKVEGELGSGVTIK---PLP-EMHVRIYLCREMAGSGTGTVVSTTFIKAEYKGRV 76
 Db 65 NITFPENGTGFVNTIAQLSDSGRYKCGLGINS-RGLSFDVSLV 109
 QY 77 TLKQYPRKNFLVEVTQLTESDGVYACGAGNMDRGTKTKVILNV 122

RESULT 6
 ID W03179 standard; Protein; 757 AA.
 AC W03179;
 DE 24-FEB-1997 (first entry)
 KW Bovine poly-immunoglobulin receptor.
 KW heavy chain; antigen binding domain; protection protein; mutants;
 KW mucosal; environment; gastrointestinal; passive; immunisation;
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
 KW poly; sorbinus; cow.
 OS Bos taurus.
 FH Key
 FT region Location/Qualifiers
 FT 13..45 /note="putative immunoglobulin binding residues
 FT of domain I"
 FT 1..120 /label=domain_I
 FT 110..230 /label=domain_II
 FT 210..340 /label=domain_III
 FT 320..450 /label=domain_IV
 FT 440..550 /label=domain_V
 FT 530..606 /note="external portions of domain VI"
 FT 550..627 /note="external portions of domain VI"
 FT 625..660 /label=transmembrane_segment
 FT 650..757 /label=intracellular_portion
 PN W09621012-A1.
 PD 11-JUL-1996.
 PR 27-DEC-1995; U16889.
 PR 30-DEC-1994; US-367395.
 PR 04-MAY-1995; US-434000.
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.
 PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
 PI Hiatt AC, Lehner T, Ma JKC;
 DR WPI: 96-333987/33.
 DR N-PSDB; T31289.
 PT Immunoglobulin and protection protein complex and its prodn. in
 PT esp. against S. mutans and S. sorbinus to prevent dental caries
 PT Disclosure; Pages 111-115; 152pp; English.
 PS The present sequence is the bovine poly-immunoglobulin (I9)
 CC receptor, a portion of which corresp. to residues 1-627, pref.
 CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
 CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
 CC The Ig of the invention comprises a pp as above in association with
 CC an Ig derived heavy chain, having at least a portion of an antigen
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
 CC gastrointestinal, environments, therefore enhancing its
 CC effectiveness in passively immunising animals against mucosal
 CC pathogens. The Ag binding domain is specifically derived from the
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
 CC S. sorbinus serotypes d and g.
 SQ Sequence 757 AA;

Query Match 8.7%; Score 245; DB 1; Length 757;
 Best Local Similarity 36.7%; Pred. No. 9.49e-09;
 Matches 40; Conservative 25; Mismatches 37; Indels 7; Gaps 5;

Db 25 PE-EVSGEGRSVSIKCYPTSVNRHRTKYWCQ-RQAQGRCTTLISSEGVSDVYGRA 82
 QY 21 PEVKVEGELGSGVTIK---PLP-EMHVRIYLCREMAGSGTGTVVSTTFIKAEYKGRV 76
 Db 83 NITFPESGTGFVVDISHLTHKDSGRYKCGLGISS-RGLNFDVSLVSD 130
 QY 77 TLKQYPRKNFLVEVTQLTESDGVYACGAGNMDRGTKTKVILNVHSE 125

RESULT 7
 ID W03181 standard; Protein; 769 AA.
 AC W03181;
 DE 24-FEB-1997 (first entry)
 KW Rat poly-immunoglobulin receptor.
 KW heavy chain; antigen binding domain; protection protein; mutants;
 KW mucosal; environment; gastrointestinal; passive; immunisation;
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
 KW poly; sorbinus.
 OS Rattus rattus.
 FH Key
 FT region Location/Qualifiers
 FT 13..45 /note="putative immunoglobulin binding residues
 FT of domain I"
 FT 1..120 /label=domain_I
 FT 110..230 /label=domain_II
 FT 210..340 /label=domain_III
 FT 320..450 /label=domain_IV
 FT 440..550 /label=domain_V
 FT 550..606 /note="external portions of domain VI"
 FT 550..627 /note="external portions of domain VI"
 FT 625..660 /label=transmembrane_segment
 FT 650..769 /label=intracellular_portion
 PN W09621012-A1.
 PD 11-JUL-1996.
 PR 27-DEC-1995; U16889.
 PR 30-DEC-1994; US-367395.
 PR 04-MAY-1995; US-434000.
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.
 PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
 PI Hiatt AC, Lehner T, Ma JKC;
 DR WPI: 96-333987/33.
 DR N-PSDB; T31291.
 PT Immunoglobulin and protection protein complex and its prodn. in
 PT plants - useful for passive immunisation against mucosal antigens,
 PT esp. against S. mutans and S. sorbinus to prevent dental caries
 PS Disclosure; Pages 123-127; 152pp; English.
 CC The present sequence is the rat poly-immunoglobulin (I9)
 CC receptor, a portion of which corresp. to residues 1-627, pref.
 CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
 CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
 CC The Ig of the invention comprises a pp as above in association with
 CC an Ig derived heavy chain, having at least a portion of an antigen
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
 CC gastrointestinal, environments, therefore enhancing its
 CC effectiveness in passively immunising animals against mucosal
 CC pathogens. The Ag binding domain is specifically derived from the
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
 CC S. sorbinus serotypes d and g.
 SQ Sequence 769 AA;

Query Match 8.7%; Score 246; DB 1; Length 769;

Best Local Similarity 35.0%; Pred. No. 8.01e-09; Matches 41; Conservative 30; Mismatches 40; Indels 6; Gaps 5;

Db 13 FSGVSTQPIPCQDQVSSIEGNSVINCYYPDTSVNRHRYKWCRO-GANGYCATLISSN 71
 QY 10 FLFVSGALRIPEVKVBEGLGGVTKC--P-LP-EHVRILYLCREMAGSGTGGVWSTT 65
 Db 72 GYLKEYSGRASLINFENSTFVNIHLQEDTGYKCGLG-TNRRGLFFDVSLEY 127
 QY 66 NFIAKEYGRVTLQYPRKNLFLVEVTLQTESDGSVYACGAGMNTDRGRTQKVTLVN 122

RESULT 8

ID W50033 standard; Protein; 532 AA.
 AC W50033; DT 26-JUN-1998 (first entry)
 DE Human immunity related factor.
 KW Lymph node; human; immunity related factor; research; treatment;
 KW Immune disease; infectious disease.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..16
 FT Peptide /label= sig_peptide
 FT Peptide 17..532
 FT Peptide /label= mat_peptide
 FT J10072495-A.
 PN 17-MAR-1998.
 PD 17-JUN-1997; 153218.
 PR 13-JUN-1996; JP-152362.
 PA (ASAH) ASAH KASEI KOGYO KK.
 DR WPI: 98-234766/21.
 DR N-PSDB; V20383.
 PT Immunity related factor - useful in the treatment of immune related
 PT and infectious diseases
 PS Claims 1, 3 and 4; pages 18-20; 21pp; Japanese.
 CC The present sequence is a lymph node derived human immunity
 CC related factor, which can be used to research and treat immune and
 CC infectious diseases.
 CC See also 532 AA.
 SQ Sequence 532 AA;

Query Match 8.0%; Score 226; DB 1; Length 532;

Best Local Similarity 39.6%; Pred. No. 2.32e-07; Matches 36; Conservative 17; Mismatches 34; Indels 4; Gaps 3;

Db 70 VSEPEGAVTIQCHYAPSVNRHRYKWCRLGPPRWICQITIVSTNYTHRYDRVALTD 129
 QY 25 VEGELGGVTKCPL-PE-M--HVRILYLCREMAGSGTGGVWSTTFFIKAEYKGRVTLKQ 80
 Db 130 FPGRLFVVRSLQSPDDIGCYLGGIGSENN 160
 QY 81 YPRKNLFLVEVTLQTESDGSVYACGAGMNTD 111

RESULT 9

ID R14670 standard; Protein; 584 AA.
 AC R14670.
 DT 30-JAN-1992 (first entry)
 DE Truncated poly Ig-receptor encoded by allele no. 1.
 KW Rabbit; insemination; pregnancy.
 OS Oryctolagus cuniculus.
 FH Key Location/Qualifiers
 FT Peptide 1..18
 FT domain 10..118
 FT domain /label= signal sequence
 FT domain /number= I
 FT domain /note= "poly-Ig binding"
 FT domain 119..223
 FT domain /number= II
 FT domain 224..332
 FT domain /number= III
 FT domain 333..441
 FT domain /number= IV

FT domain 442..552
 FT /number= V
 FT domain 553..584
 FT /number= IV
 FT /note= "incomplete"
 FN W09116061-A.
 PD 31-OCT-1991.
 PF 16-APR-1991; U02604.
 PR 16-APR-1990; US-510161.
 PA (HARD) HARVARD COLLEGE.
 PA (SURE-) INST SUISSE RECH EXPR C.
 PI Kraehenbuhl JP, Weltzin RA, Neutra MR;
 DR WPI: 91-339549/46.
 DR N-PSDB; Q14498.
 PT Stabilised poly-Ig complex contg. portion of poly-Ig receptor -
 PT useful in protection against pathogens or against pregnancy
 PS Disclosure; Fig 7; 51 pp; English.
 CC The sequence was deduced from a cDNA clone of allele no. 1 and
 CC is a truncated poly-Ig receptor. The native gene (Mostov et al)
 CC is mutated to delete the portion encoding the transmembrane and
 CC intra-cellular domains. The recombinant protein produced by
 CC expression of the sequence is used as a stabiliser protein with a
 CC compsn. can be administered directly to the mucosal surfaces of a
 CC mammal to protect against a pathogen or against insemination. It
 CC protects against allergens that contact the respiratory or digestive
 CC mucosal surfaces and protects against pregnancy by cross-linking
 CC sperm in the vagina.
 CC See also R14671.
 SQ Sequence 584 AA;

Query Match 7.9%; Score 223; DB 1; Length 584;

Best Local Similarity 45.5%; Pred. No. 3.83e-07; Matches 40; Conservative 11; Mismatches 30; Indels 7; Gaps 5;

Db 39 GDSVSICTYPTTSVTRHSRKFMCRE-EESGRCVTLAST-GYTSEYSGRGLTDFPDKG 96
 QY 30 GGSVITKC--PLPEM--HVRILYLCREMAGSGTGGVWSTTFFIKAEYKGRVTLKQYPRKN 85
 Db 97 EFVTVTDLTIONDSGSYKCGVGN-GRG 123
 QY 86 LFLVEVTQLTESDGSVYACGAGMNTDRG 113

RESULT 10

ID W03177 standard; Protein; 773 AA.
 AC W03177;
 DT 24-FEB-1997 (first entry)
 DE Rabbit poly-immunoglobulin receptor.
 KW Rabbit; immunoglobulin; receptor; protection protein; mutants;
 KW heavy chain; antigen binding domain; protection; pathogen;
 KW mucosal; environment; gastrointestinal; passive; immunisation;
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
 KW poly; sorbinus.
 OS Oryctolagus cuniculus.
 FH Key Location/Qualifiers
 FT region 21..43
 FT /note= "immunoglobulin binding residues of domain I"
 FT domain 1..118
 FT domain /label= domain_I
 FT domain 119..223
 FT domain /label= domain_II
 FT domain 224..332
 FT domain /label= domain_III
 FT domain 333..441
 FT domain /label= domain_IV
 FT domain 442..552
 FT domain /label= domain_V
 FT region 553..606
 FT /note= "external portions of domain VI"
 FT region 553..627
 FT /note= "external portions of domain VI"

FT region 630..652
 FT /label= transmembrane_segment
 FT 653..755
 FT /label= intracellular_portion
 PN WO9621012-A1.
 PD 11-JUL-1996.
 PF 27-DEC-1995; US-367395.
 PR 30-DEC-1994; US-434000.
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.
 PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
 PA (PLAN-) PLANET BIOTECHNOLOGY INC.
 PI Blatt AC, Lehner T, Ma JKC;
 DR WPI: 96-333987/33.
 DR N-PSDB: T31287.
 PT Immunoglobulin and protection protein complex and its prodn. in
 PT plants - useful for passive immunisation against mucosal antigens,
 PT esp. against S. mutans and S. sorbinus to prevent dental caries
 PS Claim 10; Pages 98-102; 152pp; English.
 CC The present sequence is the rabbit poly-immunoglobulin (Ig)
 CC receptor, a portion of which corresp. to residues 1-627, pref.
 CC 1-606, or esp. residues 21-43, 1-118, 119-223, 224-332, 333-441,
 CC 442-552, 553-606 or 553-627 comprises a protection protein (PP).
 CC The Ig of the invention comprises a PP as above in association with
 CC an Ig derived heavy chain having at least a portion of an antigen
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
 CC gastrointestinal, environments, therefore enhancing its
 CC effectiveness in passively immunising animals against mucosal
 CC pathogens. The Ag binding domain is specifically derived from the
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
 CC S. sorbinus serotypes d and g.
 SQ Sequence 773 AA;
 Query Match 7.9%; Score 223; DB 1; Length 773;
 Best Local Similarity 45.5%; Pred. No. 3.83e-07;
 Matches 40; Conservative 11; Mismatches 30; Indels 7; Gaps 5;
 Db 39 GDSVITCYPTTSVNRHSEKFCRE--EESGRGVTLAST-GYTSQYSGRGLDFFDPRG 96
 Qy 30 GGSVTIKC--PLPEN--HVRIYLCREMGAGSGTCGTVTSTNFIKAEYKGRVTLQYPRKN 85
 Db 97 EFVTVDTQNDSDSGYKCGVGN-GRG 123
 Qy 86 LFLVEVTQLTESDSGYACGAGMNTDRG 113
 RESULT 11
 ID R14671 standard; Protein; 584 AA.
 AC R14671.
 DT 30-JAN-1992 (first entry)
 DE Truncated poly Ig-receptor encoded by allele no. 2.
 KW Rabbit; insemination; pregnancy.
 OS Oryctolagus cuniculus.
 FH Key Location/Qualifiers
 FT peptide 1..118
 FT domain 10..118
 FT /label= signal sequence
 FT /number= I
 FT domain 119..223
 FT /number= II
 FT domain 224..332
 FT /number= III
 FT domain 333..441
 FT /number= IV
 FT domain 442..552
 FT /number= V
 FT domain 553..584
 FT /number= IV
 FT /note= "incomplete"
 PN WO9116061-A.
 PD 31-OCT-1991.

PF 16-APR-1991; U02604.
 PR 16-APR-1990; US-510161.
 PA (HARD) HARVARD COLLEGE.
 PA (SURE-) INST SUISSE RECH EXPER C.
 PI Kraehenbuhl JP, Weltzin RA, Neutra MR;
 DR WPI: 91-339549/46.
 DR N-PSDB: Q14499.
 PT Stabilised poly-Ig complex contg. portion of poly-Ig receptor -
 PT useful in protection against pathogens or against pregnancy
 PS Disclosure: Fig. 7; 51 pp; English.
 CC The sequence was deduced from a cDNA clone of allele no. 2 and
 CC is a truncated poly-Ig receptor. The native gene (Mostov et al)
 CC is mutated to delete the portion encoding the transmembrane and
 CC intra-cellular domains. The recombinant protein produced by
 CC expression of the sequence is used as a stabiliser protein with a
 CC poly-Ig specific for a selected antigen or family of antigens. The
 CC compsn. can be administered directly to the mucosal surfaces of a
 CC mammal to protect against a pathogen or against insemination. It
 CC protects against allergens that contact the respiratory or digestive
 CC mucosal surfaces and protects against pregnancy by cross-linking
 CC sperm in the vagina.
 CC See also R14670.
 SQ Sequence 584 AA;
 Query Match 7.4%; Score 209; DB 1; Length 584;
 Best Local Similarity 43.2%; Pred. No. 3.32e-06;
 Matches 38; Conservative 13; Mismatches 30; Indels 7; Gaps 5;
 Db 39 GDSVITCYPTTSVNRHSEKFCRE--EESGRGVTLAST-GYTSQYSGRGLDFFDPRG 96
 Qy 30 GGSVTIKC--PLPEN--HVRIYLCREMGAGSGTCGTVTSTNFIKAEYKGRVTLQYPRKN 85
 Db 97 EFVTVNLRLSDSGYKCGVGN-GRG 123
 Qy 86 LFLVEVTQLTESDSGYACGAGMNTDRG 113
 RESULT 12
 ID Y05069 standard; Protein; 205 AA.
 AC Y05069;
 DT 16-JUN-1999 (first entry)
 DE Human PIGR-2 protein sequence.
 KW PIGR-2; human; autoimmune disease; rheumatoid arthritis; psoriasis;
 KW Multiple Sclerosis; Systemic Lupus Erythematosus; diagnosis; therapy;
 KW Inflammatory Bowel Disease.
 OS Homo sapiens.
 PN EP-905237-A2.
 PD 31-MAR-1999.
 PF 07-AUG-1998; 306323.
 PR 21-NOV-1997; US-976293.
 PR 23-AUG-1997; US-056774.
 PA (SMIR) SMITHKLINE BEECHAM CORP.
 PI Sweet RW, Truneh A, Wu S;
 DR WPI: 99-192665/17.
 DR N-PSDB: X28250.
 PT New polypeptides encoding human PIGR-2 useful for treating diseases
 PT such as rheumatoid arthritis and multiple sclerosis
 PS Claim 11; Page 17; 23pp; English.
 CC This sequence is the human PIGR-2 protein of the invention.
 CC Autoimmune diseases involving altered expression or activity of PIGR-2
 CC may include rheumatoid arthritis, Multiple Sclerosis, psoriasis,
 CC Systemic Lupus Erythematosus and Inflammatory Bowel Disease. These
 CC diseases can be diagnosed or susceptibility to them predicted by:
 CC (1) determining whether there is a mutation in the genomic copy of the
 CC gene encoding PIGR-2; or (2) measuring the amount of PIGR-2 in a sample
 CC derived from the patient. Patients deficient in PIGR-2 can be treated by
 CC administering either the PIGR-2 DNA or its complement or an agonist of
 CC PIGR-2 to the patient. Patients with excessive expression or activity of
 CC PIGR-2 can be treated by administering an antagonist of PIGR-2, an
 CC antisense nucleic acid molecule which inhibits the expression of PIGR-2
 CC or administering sufficient PIGR-2 to compete with the endogenous
 CC activity. PIGR-2 can be used to identify its agonists by contacting a
 CC cell expressing PIGR-2 with a candidate compound in the presence of a

Tue Sep 12 08:57:57 2000

OM of: US-09-135-238B-2 to: N_Geneseq_36.* out_format : pfs
Date: Sep 12, 2000 5:36 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-DB=N_Geneseq_36 -OFMT=fastap -SUFRIX=p2n.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCALLGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=1000000
-USER=US09135238 -CGN1_1_123 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:
Query: US-09-135-238B-2
Query length: 390
Database: N_Geneseq_36.*
Database sequences: 311585
Database length: 125096042
Search time (sec): 122.100000

score_list:	Sequence	Strd Orig	zScore	EScore Len	Documentation
	N_Geneseq_36:X28178	+ 2047.00	2129.34	5.3e-111	2040
	N_Geneseq_36:X28179	+ 1125.00	1174.09	8.5e-58	1047
	N_Geneseq_36:T31290	+ 198.00	199.21	0.0017	3095
	N_Geneseq_36:T31291	+ 188.00	170.27	0.0697	29392
	N_Geneseq_36:T31292	+ 187.50	192.24	0.0042	1911
	N_Geneseq_36:T31293	+ 185.50	185.74	0.0096	3269
	N_Geneseq_36:T31294	+ 185.00	189.53	0.0059	1936
	N_Geneseq_36:T31295	+ 182.00	186.82	0.0083	1839
	N_Geneseq_36:T31296	+ 182.00	186.01	0.0093	2031
	N_Geneseq_36:T31297	+ 182.00	183.23	0.0136	2919
	N_Geneseq_36:T31298	+ 182.00	181.02	0.0171	3630
	N_Geneseq_36:T31299	+ 171.00	174.33	0.0414	2084
	N_Geneseq_36:T31300	+ 171.00	172.62	0.0515	2586
	N_Geneseq_36:T31301	+ 166.50	170.51	0.0676	1879
	N_Geneseq_36:T31302	+ 166.50	165.34	0.1312	3517
	N_Geneseq_36:T31303	+ 159.50	167.34	0.1015	1137
	N_Geneseq_36:T31304	+ 145.00	148.11	1.20	1876
	N_Geneseq_36:T31305	+ 143.50	133.93	7.37	8690
	N_Geneseq_36:T31306	+ 142.00	130.90	0.8361	914
	N_Geneseq_36:T31307	+ 139.50	130.12	12.02	8321
	N_Geneseq_36:T31308	+ 138.50	146.63	1.44	985
	N_Geneseq_36:T31309	+ 138.50	146.63	1.44	985
	N_Geneseq_36:T31310	+ 136.50	149.12	1.05	565
	N_Geneseq_36:T31311	+ 136.50	140.14	3.32	1683
	N_Geneseq_36:T31312	+ 136.00	123.79	27.06	11529
	N_Geneseq_36:T31313	+ 135.00	123.35	28.60	10704
	N_Geneseq_36:T31314	+ 131.00	118.16	55.71	12132
	N_Geneseq_36:T31315	+ 130.00	133.70	7.59	1617
	N_Geneseq_36:T31316	+ 129.50	134.18	7.13	1431
	N_Geneseq_36:T31317	+ 129.50	134.18	7.13	1431
	N_Geneseq_36:T31318	+ 127.00	128.05	15.66	2197
	N_Geneseq_36:T31319	+ 126.50	120.88	39.26	4926
	N_Geneseq_36:T31320	+ 126.50	114.03	94.63	11336
	N_Geneseq_36:T31321	+ 125.50	129.22	13.48	1576
	N_Geneseq_36:T31322	+ 125.50	122.89	30.35	3400
	N_Geneseq_36:T31323	+ 125.50	116.64	67.64	7266
	N_Geneseq_36:T31324	+ 124.50	116.45	69.37	6557
	N_Geneseq_36:T31325	+ 124.50	108.70	187.39	16812
	N_Geneseq_36:T31326	+ 124.00	125.90	20.63	1951
	N_Geneseq_36:T31327	+ 124.00	124.32	25.27	2364

N_Geneseq_36:T50962	+	124.00	115.30	80.34	7073
N_Geneseq_36:T50963	+	123.50	125.47	21.80	1929
N_Geneseq_36:T50964	+	123.00	125.54	21.60	1795
N_Geneseq_36:T50965	+	123.00	113.19	105.27	8051
N_Geneseq_36:T50966	+	123.00	113.08	106.78	8160

seq_name: N_Geneseq_36:X28178
seq_documentation_block:
ID X28178 standard: cDNA; 2040 BP.
AC X28178;
DE 16-JUN-1999 (first entry)
DT Human PIGRL-1 coding sequence.
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;
KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;
KW diagnosis; therapy; ss.
OS Homo sapiens.
PN EP-905238-A2.
PD 31-MAR-1999.
PF 14-AUG-1998; 306487.
PR 30-OCT-1997; US-961564.
PR 25-AUG-1997; US-056935.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Sweet RW, Truneh A, Wu S;
DR WPI: 99-192666/17.
DR P-PSDB: Y05001.
PT New polypeptides encoding PIGRL-1 useful for treating diseases such
PT as X-linked Severe Combined Immunodeficiency
PS Claim 4; Page 18-19; 26pp; English.
CC This sequence encodes the human PIGRL-1 protein of the invention.
CC Autoimmune diseases involving altered expression or activity of PIGRL-1
CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined
CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be
CC diagnosed or susceptibility to them predicted by: (1) determining whether
CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or
CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.
CC Patients deficient in PIGRL-1 can be treated by administering either the
CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.
CC Patients with excessive expression or activity of PIGRL-1 can be treated
CC by administering an antagonist of PIGRL-1, an antisense nucleic acid
CC molecule which inhibits the expression of PIGRL-1 or administering
CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can
CC be used to identify its agonists by contacting a cell expressing PIGRL-1
CC with a candidate compound in the presence of a signal system and noting
CC the candidate as an agonist if a signal is produced. The same method can
CC be used to identify antagonists of PIGRL-1 but the presence of an
CC antagonist is indicated by a decrease in production of the signal.
CC Antibodies against PIGRL-1 may be used to isolate or identify clones
CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to
CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.
CC This information may then be correlated with the incidence of autoimmune
CC disease in those patients to identify whether the mutation causes the
CC disease.
SQ Sequence 2040 BP; 478 A; 616 C; 494 G; 452 T;
alignment_scores:
Quality: 2047.00 Length: 390
Ratio: 5.262 Gaps: 0
Percent Similarity: 99.744 Percent Identity: 99.744
alignment_block:
US-09-135-238B-2 x X28178 ..
Align seg 1/1 to: X28178 from: 1 to: 2040
1 MetaspargTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLe 17
146 ATGGACTTCCTGGCTTTGGCCACCTTACTTCTCGCAGTATCAGGGGCCCT 195
17 uatgIleLeuProGluValLysValGluGlyGluLeuGlyGlySerValt 34
196 GAGGATCTCTCCAGAGTAAGGTAGAGGGGAGCTGGGGGGATCAGTTA 245

```

34  hrIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 50
|||||
246 CCATCAAGTGCCTTCCTGAAATGCATGTGAGATATATCTGTGCGG 295
51  GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67
|||||
296 GAGATGGCTGGATCTGGAACATGTGTACCGTGGTATCCACCACTT 345
67  eIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgL 84
|||||
346 CATCAGGCGAGAAATACAGGGCGGAGTTACTCTGAGCAATACCCAGCA 395
84  ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
|||||
396 AGAATCTGTCTTCTAGTGGAGGTAAACACAGCTGCAGAAAAGTGACAGCGGA 445
101 ValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117
|||||
446 GTCTATGCTCGGAGCGGCATGAACACAGACCGGGGAAAGACCCAGAA 495
117 sValThrLeuAsnValHisSerGluTyrGluProSerTrpGluGlnP 134
|||||
496 AGTCACCTGTAATCTCCACAGTGAATACGAGCCATCATGGGAAGCAGC 545
134 roMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMet 150
|||||
546 CAATGCCTGAGACTCCAAATGGTTTCATCTGCCCTATTGTTCAGATG 595
151 ProAlaTyrAlaSerSerLysPheValThrArgValThrThrProAl 167
|||||
596 CTGTCATATGCCAGTCTTCCAAATTCGTACACAGAGTTACACACACAGC 645
167 aGlnArgGlyLysValProProValHisHisSerSerProThrThrGlnI 184
|||||
646 TCNAGGGGCGAGGTCTCTCCAGTTCACCACTCTCTCCCGACCCAAA 695
184 leThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLys 200
|||||
696 TCACCCAGCGCCCTCGAGTGTCCAGAGCATCTTCAGTAGCAGGTGACAAG 745
201 ProArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuG 217
|||||
746 CCCCAGACCTTCCTGCATCCACTACAGCTCAAAATCTCAGCTCTGGA 795
217 uGlyLeuLeuLysProGlnThrProSerTyrAsnHisHisThrArgLeuH 234
|||||
796 GGGGCTGTCTAAGCCCGACAGCCAGCTACAACACACACACAGGCTGC 845
234 isArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGln 250
|||||
846 ACAGCAGAGACACTGGACTATGCTACAGTCTGGAGGAGGCGCAA 895
251 GlyPheHisIleLeuLeuProThrIleLeuGlyLeuPheLeuAlaLe 267
|||||
896 GGATTTCATCTGATCCCGACCATCTCGGCCCTTTCTGCTGGCACT 945
267 uLeuGlyLeuValValLysArgAlaValGluArgArgLysAlaLeuSerA 284
|||||
946 TCTGGGGCTGGTGTGAAAGGGCGGTGAAAGGGGAAAGCCCTCTCCA 995
284 rArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln 300
|||||
996 GCGCGGCCCGCGACTGGCGGTGAGGATGCGCGCTTCGGAGAGCTCCAG 1045
301 ArgProArgGlySerProArgProArgSerGlnAsnAsnIleTyrSerAl 317
|||||
1046 AGCCCCCGGGTGGCGCGAGCGCGCTCCCAAAACAAACATCTACAGCG 1095
317 aCysProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlap 334
|||||
1096 CTGCCCCGCGCGCTCTGTGGAGGCGAGCTGCAGGCACAGGGGAGGCC 1145

```

```

334  toValProGlyProGlyAlaProLeuProAlaProLeuGlnValSer 350
|||||
1146 CGGTTCCCGCCCGGAGCGCGTGTCCCGCCCGCCCGCTCGAGGTGCT 1195
351 GluSerProTrpLeuHisAlaProSerLeuLysThrSerCysGluTyrVa 367
|||||
1196 GAATCTCCTCGCTCATGCGCCCATCTCTGAAGACACAGCTGTGAATACGT 1245
367 lserLeuTyrHisGlnProAlaAlaMetMetGluAspSerAspSerAspA 384
|||||
1246 GAGCCTCTACCAACAGCCTCGCCCATGATGAGGACAGATGATTTCAGATG 1295
384 spTyrIleAsnValProAla 390
|||||
1296 ACTACATCAATGTTCTGCC 1315

```

seq_name: N_Geneseq_36:X28179

seq_documentation_block:

ID X28179 standard; cDNA; 1047 BP.

AC X28179;

DT 16-JUN-1999 (first entry)

DE Human PIGRL-1 coding sequence fragment.

KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;

KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;

KW diagnosis; therapy; ss.

OS Homo sapiens.

PN EP-905238-A2.

PD 31-MAR-1999.

PF 14-AUG-1998; 306487.

PR 30-OCT-1997; US-961564.

PR 25-AUG-1997; US-056935.

PA (SWIK) SMITHKLINE BEECHAM CORP.

PI Sweet RW, Truneh A, Wu S;

DR WPI; 99-192666/17.

DR P-PSDB; Y05002.

PT New polypeptides encoding PIGRL-1 useful for treating diseases such

as X-linked Severe Combined Immunodeficiency

PT Disclosure; Page 8; 26pp; English.

PS This sequence encodes the human PIGRL-1 protein of the invention.

CC Autoimmune diseases involving altered expression or activity of PIGRL-1

may include hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined

Immunodeficiency (XSCID) and IgA deficiency. These diseases can be

diagnosed or susceptibility to them predicted by: (1) determining whether

there is a mutation in the genomic copy of the gene encoding PIGRL-1; or

(2) measuring the amount of PIGRL-1 in a sample derived from the patient.

Patients deficient in PIGRL-1 can be treated by administering either the

PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.

Patients with excessive expression or activity of PIGRL-1 can be treated

by administering an antagonist of PIGRL-1, an antisense nucleic acid

molecule which inhibits the expression of PIGRL-1 or administering

sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can

be used to identify its agonists by contacting a cell expressing PIGRL-1

with a candidate compound in the presence of a signal system and noting

the candidate as an agonist if a signal is produced. The same method can

be used to identify antagonists of PIGRL-1 but the presence of an

antagonist is indicated by a decrease in production of the signal.

Antibodies against PIGRL-1 may be used to isolate or identify clones

expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to

identify chromosomal mutations in the gene encoding PIGRL-1 in patients.

This information may then be correlated with the incidence of autoimmune

disease in those patients to identify whether the mutation causes the

CC disease.

SQ Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T;

alignment_scores:

Quality: 1125.00

Ratio: 4.429

Percent Similarity: 92.364

Percent Identity: 90.182

alignment_block:

US-09-135-238B-2 x X28179 ..

Align seg 1/1 to: X28179 from: 1 to: 1047

```

1 MetAspArgTyrPheLeuTyrPheLeuProValSerGlyAlaLe 17
150 ATGGACTTCTGGCTTGGCCACTTACTTCTGCGCAGTATCAGGGCCCT 199
17 uArgileLeuProGluValValGlyValGluGlyGlySerValT 34
200 GAGGATCTCCAGAGTAAGGTAGAGGGGAGCTGGCGGATCAGTTA 249
34 hrLeLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 50
250 CCATCAAGTGCCTCTCTGAAATGATGTGAGGATATATCTGTGCGG 299
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67
300 GAGATGGCTGGATCTGGACATGTGGTACCTGGTATCCACCACTT 349
67 elLeLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgL 84
350 CATCAAGGCAGAAATCAAGGGCCGAGTTACTCTGAAGCAATACCCAGCA 399
84 ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
400 AGAATCTGTCTAGTGGAGGTAAACAGCTGACAGAAAGTGACAGCGGA 449
101 ValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117
450 GTCTATGCTCGGGAGC.GGCATGAACACAGACCCGGGAAAGACCCAGAA 498
117 svAlThrLeuAsnValHisSerGlyTyrGluProSerTrpGluGlnP 134
499 AGTCACCTTGAATGCCAGAGTAATACGACCATCATGGAGAGAGC 548
134 roMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMet 150
549 CAATGCTCAGACTCCAAATGTTTCTATCTGCCTATTTGTTCCAGATG 598
151 ProAlaTyrAlaSerSerSerLysPheValThr...ArgValThrPro 166
599 CTTGCATATGCGGCTTCTCCACATTCGTAAACCGCAGAGTTACACACCA 648
167 Ala.GlnArgGlyLysValProValHisHisSerSerProThrThrG 183
649 GCTTCAAGGGGCAAGTCCCTCCAGTTCACACTCTCTCCCCACACCC 698
183 lnIle...ThrHisArgPro.ArgValSerArgAlaSerSerValAlaGlyA 199
699 AAATTCAACCCCGCCCTTCGAGTGTNCAGAGCATCTTCAGTAGCAGGTG 748
199 sPlyProArg...ThrPheLeuProSerThrThrAlaSerLysIleSerAl 215
749 ACAAGCCCGAAATTTCTGTCATCCACTACAGCTCAAAATCTCAGC 798
215 aLeuGluGlyLeuLeuLys...ProGlnThrProSerTyrAsnHisIst 231
799 TCTGGAAGGCTGCTTCAAGCCCAAGAGCCGCCAGCTTACAACAN.CACA 847
231 hrArgLeuHisArgGlnArgAlaLeuAsp...Tyr.GlySerGlnSerGl 246
848 CCAGGCTGCACAGGACAGACACTGGATCTATATGGGNTCACAGTCTGG 897
246 yArgGluGly.....GlnGlyPhe...HisIleLeuIlePro...ThrIle 259
898 G...GAGGGAANGNCNCAAGGATTTTNCATCTCTCTGATTCCTCCGACCATC 944
260 LeuGlyLeuPhe 263
945 NTGCGGCTTT 956

```

seq_name: N_Geneseq_36.T31290

seq_documentation_block:

```

ID T31290 standard; CDNA; 3095 BP.
AC T31290;
DE 24-FEB-1997 (first entry)
KW Mouse poly-immunoglobulin receptor, cDNA.
KW Mouse; immunoglobulin; receptor; protection protein; mutans;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; murine; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT cds 85..2400
FT cds /*tag= a
PN WO9621012-A1.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI: 96-333987/33.
DR P-PSDB: W03180.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - usefui for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure: Pages 117-121; 152pp; English.
CC The present sequence encodes the mouse poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606 or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 3095 BP; 861 A; 796 C; 784 G; 654 T;

alignment_scores:
Quality: 198.00 Length: 168
Ratio: 1.980 Caps: 5
Percent Similarity: 59.524 Percent Identity: 31.548

alignment_block:
US-09-135-238B-2 x T31290 ..
Align seg 1/1 to: T31290 from: 1 to: 3095

10 PheLeuProValSerGlyAlaLeuArgIleLeuProGluValLysValG1 26
121 TTTTACGGGTCTCCACAAAAGCCCATATTTGTCCTCCCGAGGAGTGAG 170
26 uGlyGluLeuGlySerValThrIleLysCysProLeuProGlu... 41
171 TAGTATAGAAGCGGACTGTGTTCCATCAGGTGCTACTACCCAGACACCT 220
42 .....MetHisValArgIleTyrLeuCysArgGluMetAlaGlySer 55
221 CTGTCAACCGGCACACCGGAAATACTGTGCGCAAGAGGCC...AGC 267
56 GlyThrCysGlyThrValValSerThrThrAsnPheIleLysAlaGluT 72
268 GGCATGTGCACACGCTCATCTCTTCAATGGCTACCTCTCCAGGAGTA 317
72 rLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeu 89

```

```

318 TTCAGCAGAGCAACCTCATCACTCCAGAGAACACACATTTGTGA 367
89 algluValThrGlnLeuThrGluSerAspSerGlyValThrAlaCysGly 105
   ::::: ||||| ::::: ||||| ::::: |||||
368 TTAACATGTAGCAGCTCACCAGGACGACACTGGAGCTACAAGTGTGGC 417
   ||| ::::: ||||| ::::: |||||
106 AlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVa 122
   ||| ::::: ||||| ::::: |||||
418 CTGGGT...ACAGTAACCGAGGCTGCTTCATGTCAGCTGAGGT 464
122 lHisSerGluTyrGluProSerTyrGluGluGlnProMetProGluThrP 139
   | ::::: |||||
465 C.....AGCCAGGTCTCTGAGTGC 484
139 rolystrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSer 155
   ||| ::::: |||||
485 CGAGTGACACCCAGTC.....TACACAAAG 510
156 SerSerLysPheValThrArgValThrThrProAlaGlnArgGlyLysVa 172
   :: ::::: |||||
511 GACATAGGCAGAAATGTGACCATTTGAATGCCCTTTCAAAAGGGAGAATGT 560
172 lPro 173
   ||||
561 TCCC 564

seq_name: N_Geneseq_36.v15422
seq_documentation_block:
ID V15422 standard; DNA; 29392 BP.
AC V15422;
DT 11-JUN-1998 (first entry)
DE Mouse poly Ig receptor protein gene.
KW Mouse; poly Ig receptor protein; plgR protein; pig; deficiency;
KW knockout mouse; disease model; ds.
OS Mus sp.
PN J10057066-A.
PD 03-MAR-1998.
PF 19-AUG-1996; 217154.
PR 19-AUG-1996; JP-217154.
PA (HONS.) TAKUJI HONSHA KK.
DR WPI; 98-254323/23.
PT Mouse pig receptor protein gene - used for preparing gene knockout
PT mice, useful for study of human poly Ig receptor protein deficiency
PS Claim 1; Page 4-14; 18pp; Japanese.
CC The present sequence represents the mouse poly Ig receptor protein
CC gene, which has a 29392 bp sequence. The new gene can be used to
CC produce a gene knockout mouse, useful as a disease model of human
CC poly Ig receptor protein deficiency.
SQ Sequence 29392 BP; 8318 A; 6747 C; 6514 G; 7813 T;

alignment_scores:
  Quality: 188.00      Length: 97
  Ratio: 2.648        Gaps: 3
Percent Similarity: 73.196 Percent Identity: 40.206

alignment_block:
US-09-135-238B-2 x V15422 ..
Align seg 1/1 to: V15422 from: 1 to: 29392

30 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 42
||| ||||| ::::: |||||
18245 GCGGACTCTGTTTCATACAGTGTGCTACTACCCAGACACCTCTGTCACACG 18294

42 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
||| ::::: ||||| ::::: |||||
18295 GCACACCCGGAAATACTGTGTGCGACAGGAGCC...AGCGCATGTGCA 18341

59 lYthrValIleSerThrAsnPheIleLysAlaGluTyrLysGlyArg 75
||| ::::: ||||| ::::: |||||
18342 CACCGCTCATCTCTCAATGGCTACTCTCCAGGAGTATTCAGGCAGA 18391

```

```

76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92
   ::::: ||||| ::::: ||||| ::::: |||||
18392 GCCAACCTCATCAACTTCCAGAGAACACACATTTGTGATTAAACATTGA 18441

92 rGlnLeuThrCluSerAspSerGlyValThrAlaCysGlyAlaGlyMeta 109
   ||||| ::::: ||||| ::::: |||||
18442 GCAGCTCACCCAGAGACACACTGGAGCTACAAAGTGTGCTGGGT...A 18488
   ::::: ||||| ::::: |||||
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
   ::::: ||||| ::::: |||||
18489 CCAGTAACCCAGGCTGCTTCGATGTCAGCTGAGGTGTC 18529

seq_name: N_Geneseq_36.v20383
seq_documentation_block:
ID V20383 standard; cDNA to mRNA; 1911 BP.
AC V20383;
DT 26-JUN-1998 (first entry)
DE cDNA for human immunity related factor.
KW Lymph node; human; immunity related factor; research; treatment;
KW immune disease; infectious disease; ds.
OS Homo sapiens.
FH Key
FT CDS
FT sig_peptide
   /tag= a
   /tag= b
   /tag= c
FT mat_peptide
   /product= immunity_related_factor
FT J10072495-A.
PD 17-MAR-1998.
PF 11-JUN-1997; 153218.
PR 13-JUN-1996; JP-152362.
PA (ASAH.) ASAH KASEI KOGYO KK.
DR WPI; 98-234766/21.
DR P-PSDB; W50033.
PT Immunity related factor - useful in the treatment of immune related
PT and infectious diseases
PS Claim 15; Pages 18-20; 21pp; Japanese.
CC The present sequence encodes a lymph node derived human immunity
CC related factor, which can be used to research and treat immune and
CC infectious diseases.
SQ Sequence 1911 BP; 490 A; 541 C; 525 G; 355 T;

alignment_scores:
  Quality: 187.50      Length: 575
  Ratio: 0.906        Gaps: 20
Percent Similarity: 36.000 Percent Identity: 18.957

alignment_block:
US-09-135-238B-2 x V20383 ..
Align seg 1/1 to: V20383 from: 1 to: 1911

3 ArgTyrLeuTyrProLeuTyrPheLeuProValSerGlyAlaLeuArgTil 19
||||| ::::: |||||
170 AGATGGCTGTGGGAG...GGCTCTCTCCCTCCAGGACCCATCTCCGGGC 216

19 eLeuProGluValLys..... 24
::: :::::
217 CATGGAAACACTCAGGCTTCCTCGCCCTCTGTCGGCGGAGGAGCT 266

25 .....ValcluglyVlu 28
||||| ::::: |||||
267 CCTTTGCAGCTCCAAATTCATTGAAGGCTCAAGGCTGTGTCTCAGGGAG 316
||||| ::::: |||||
29 LeuGlyGlySerValThrIleLysCysProLeuPro.....G1 41
||||| ::::: |||||
317 CTGGAGGAGCTGTACCATCGAGTGCCATTATGCCCTCATCTGTCAA 366

```


41 uMetHisValArgIleTyrLeuCysArgGluMetalaglySerGlyThrC 58
: ||| ||| ||| |||||
367 CAGGCACCAAGAGAAAGTACTGGTCCTCTGGGGCCCCAAGATGGATCT 416

58 ySgLYThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGly 74
||| ||||| ||||| ||||| :|||:
417 GCCAGACCATTGTGTCCACCAACCAAGTACTCACCATCGCTATCGTGAC 466

75 ArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluVa 91
|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
467 CGTGTGGCCCTCACAGACTTCCACAGAGAGGCTTGTGTGGTGAGGCT 516

91 lThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyM 108
||| ||||| :|||:||||| ||||| |||||
517 GTCCCAACTGTCCC GGATGACATCGCATCTACCTCTCGCGCACTGGAA 566

108 etAsnThrAspArg..... 112
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
567 GTGAACAACATCGCTTCTTAGCATGAATCTGACATCTCTCGAGGT 616

112 112
:
617 CCCGCCAGCACCTCCCCACAGCCACTCCAGCTGCTGGGGAGCTCACCAT 666

113 GlyL 114
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
667 GAGATCCTATGAACAGAGCTCTCAGTGGCCACAGATGGACCCAGGA 716

114 ysThrGln..... 116
|||||
717 CCACCCAGACCTTAGGACAGGGGACAGCATGGGACACAGTTGCTTCCACT 766

117 LysValThrAsnValHisSerGluTyrGluPro.. 128
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
767 CCAGGAACACAGCAAGACTACAGCTCAGCTGAGGGAAGACAGACCCAG 816

129 SerTrpGluGlu..... 132
:
817 AGCAACACAGCCAGCAGCTCCAGGACAGAGCAGCTGGGACAGGGTCTTG 866

133 GlnProMetProGluThrProLysTrpPheHisLeuPro 145
|||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:
867 TCAAAGCACCTGCTCCGATTCCAGAGTCCA..... 898

146 TyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysPheValThrAr 162
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
899 CCTCAAAGACAGCAAGCATGTCCAA 924

162 gValThrThrProAlaGlnArgGlyLysValProValHisHisSerS 179
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
925 TACAACAGAAGGTGTTGGAGGC..... 949

179 erProThrThrGlnIleThrHisArgProArgValSerArg..... 192
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
950 .. ACCAGAAGCTCGGTGACAAACAGGCTAGACCCAGCAAGACAGAGG 997

193 ... AlaSerSerValalaGlyAspLysProArg..... 202
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
998 GAGATGACAACACTACCAAGGCTGATAGGCCAAGGAGGACATAGAGGGGT 1047

202 202
:
1048 CAGGATAGCTCTTGATGCAGCCAAAAAGGTCCTAGGAACCATTTGGGCCAC 1097

203 ThrPheLeuProSerThrThr 209
:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1098 CAGCTCTGGTCTCAGAAACTTTGGCCTGGGAAATCTCCCAACAACAAG 1147

210 AlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProGlnThrProSe 226
::: |||:||||:||||:||||:||||:||||:||||:||||:||||:
1148 CCAGTTTCAAGCAACAATCTCAGGGGTTCATTGGAGAACCACTCCAG 1197

226 tTyraSnHisHisThrArg..... 232

```

: ::: |||
1198 TGCAGGCATGTGCACCTCTGGGAACCTCCAGCTGCAGATGTGGATCTTGG 1247
233 .....LeuHisArgGlnArgAlaLeuAspTyr 241
1248 GAATCCAGCTGCAGATGTGTGGACCAGCATGGAGCGACATCTGGGGA 1297
242 GlySerGlnSerGly..... 246
||||| :::|||
1298 GGAAGCCCTCGAGGGGACCTAGATGTGCCACTGGAGACAGAGGTCCCCA 1347
246 ..... 246
1348 AGCAACACTGAGCCAGACCCCGCAGTAGGACCTCTGGGACCCCTGGCA 1397
247 .....ArgGluGlyGlnGlyPheHis 253
1398 AGGAGTCTCCGTGAGCGTACTTTTCCAGAAGATGAAGCAGCTCTCGG 1447
254 IleLeuIlePro.....ThrIleLeuGlyLeuPheLeuAlaLeuLe 268
||| ||| :::|||:|||||:|||||
1448 ACCCTGCCTCTCTCTCTACCATGTCTGGCCCTGTTTATGCTTATGGCTCT 1497
268 uGlyLeuValValIlysArgAlaValGluArgArgLysAlaLeuSerArg 285
|||||: :::|||:|||||
1498 GGTCTATTG.....CAAAGGAAGCTCTGGAGAGGA 1529
285 rgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln... 300
||| :::||| :::|||:|||||:|||||
1530 GGACCTCTCAGGAGGAGAAAGGGTCACCTTAATTCAGATGACACATTT 1579
301 .....ArgProArgGlySerProArgProArgSerGlnAsnAsnI 314
:|||||: :::|||:|||||:|||||
1580 CTGGAAGTGAACCCCAAGCAGACAGCTGCCCATGTGGAAGAAAGAT 1629
314 eTyr.....SerAlaCysProArgArgAlaArgGlyAlaAspAlaAla 329
: :::||| :||| :||| :|||||
1630 GCTCCAGGATGACTCTCTCTCT.....GCTGGGCGCAGCC 1664
329 lyThrGlyGluAlaProValProGlyPro.....Gly 339
|||||: :|||||
1665 TGACTGCCCCAGAGAAATCCAGGACCTTGAGGACAGAGATGAACT 1714
340 AlaProLeuPro.....ProAl 345
||| :|||
1715 GCTCAGTTTACATCTGGGAGAGCAAGATCAAGGCCTTCAGGACCCCA 1764
345 aProLeuGlnValSerGluSerPro 353
|||||: :|||
1765 GCCTCTTTCATCATCTCTCTCCA 1789

seq_name: N_Genseq_36:T31291

seq_documentation_block:
ID T31291 standard; cDNA; 3269 BP.
AC T31291;
DT 24-FEB-1997 (first entry)
DE Rat poly-immunoglobulin receptor, cDNA.
KW heavy chain; antigen binding domain; protection protein; mutants;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; ss.
OS Rattus rattus.
FH Key Location/Qualifiers
FT cds 74..2383
FT /tag= a
FT
FT W09621012-Al.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.

```



```

74 yArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluV 91
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
255 CAGGGCTACCTCAACCTCCCGGAGAACGGCACATTCGTGGTGAACA 304
91 aThrGlnLeuThrGlnSerAspSerGlyValTyrAlaCysGlyAlaGly 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
305 TTGCCACGCTGAGCCAGAGTACCTCCGGGCGCTACAAGTGTGGCCTGGGC 354
108 MetAsnThr..... 110
|||||:|||||:|||||:|||||:|||||:|||||:
355 ATCAATAGCCGAGCGCTGTCTTTGTATGTAGCTGGAGGTACGCCAGG 404
111 .....AspArgGlyLysT 115
|||||:|||||:|||||:|||||:|||||:|||||:
405 TCCTGGGCTCTTAATGACACTAAGTCTACACAGTGGACCTGGCGAGAA 454
115 hrGlnLysValThrLeuAsnValHisSerGluTyrGluProSerTrpGlu 131
|||:|||||:|||||:|||||:|||||:|||||:|||||:
455 CG.....GTGACCATCACTGCCCTTTCAAGACTGAGAATGCTCAAAAG 498
132 GluGln.PrometProGlu.....ThrProLysTrpPheHisL 144
|||||:|||||:|||||:|||||:|||||:|||||:
499 AGAAGTCTGTGTACACAGATAGGCTGTACCCGTGCTGTGTATCGA 548
144 euProTyrLeuPheGlnMetProLysTyrAlaSerSerLysPheVal 160
|||||:|||||:|||||:|||||:|||||:|||||:
549 CTCAGTGTGTATGTGAATCCCACTATACAGGAAGAATACGCTTGATA 598
161 ThrArgVal..... 163
|||||:|||||:|||||:|||||:|||||:|||||:
599 TTCAGGGTACTGGCCAGTACTGTTCAGGTTGTATCAACCACTCAGG 648
164 .....ThrThrProAlaGlnArgGlyLysValProProV 175
|||||:|||||:|||||:|||||:|||||:|||||:
649 CTCAGGATGCTGGCAGTATCTCTGCCAGGCTGGGATGATTTCCAATAG 698
175 aHisHisSerSerProThrThrGlnLeuHisArgProArgValSer 191
|||:|||||:|||||:|||||:|||||:|||||:|||||:
699 TAATAGAGAAATGCTGACC..... 718
192 ArgAlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerTh 208
|||||:|||||:|||||:|||||:|||||:|||||:
719 .....TCCAAGTGTAAAGCCGAGCCGAGCTGTTTATGAAGACCTC 762
208 rThrAlaSerLysLleSerAlaLeuGluGlyLeuLysProGlnThr. 224
|||||:|||||:|||||:|||||:|||||:|||||:
763 AGGGCTCAGTACCTTCCACTGTGCCCTGGCCCTGAGTGGCAACGT 812
225 ..ProSerTyrAsnHisHisThrArgLeuHisArg..... 235
|||||:|||||:|||||:|||||:|||||:|||||:
813 GGCCAAATTTCTGCCACAGAGCAGTGGGAAACTGTGACGTGGTCG 862
236 .....GlnArgAlaLeuAspTyrGlySerGlnSe 245
|||||:|||||:|||||:|||||:|||||:|||||:
863 TCACACCTAGGGAAGAGGCCCGCCAGCTTTGAGGCGAGATCTCTGTC 912
245 r..... 246
|||||:|||||:|||||:|||||:|||||:|||||:
913 AACCCCCAGGACAGGCTGCTATTGAGTGTGTATCAGAGCCCTGAG 962
246 lYArgGluGlyGlnGlyPheHisLleLeuIleProThrIleLeuGlyLeu 262
|||||:|||||:|||||:|||||:|||||:|||||:
963 GAAGGAGGATGCGAGGGCGCTACCTGTGTGGAGCCCAATT..... 1000
263 PheLeuLeuAlaLeuGlyLeuValLysArgAlaValGluArgAr 279
|||||:|||||:|||||:|||||:|||||:|||||:
1001 .....CGGATGCTCAGCTGCAG 1017
279 gLysAlaLeuSerArgAlaArgArgLeuAlaValArgMetArgAlaL 296
|||||:|||||:|||||:|||||:|||||:|||||:
1018 GAAGGCT.....CGCCTATCCAGGC 1037

```

```

296 euGluSerSerGlnArgProArgGlySerProArgProArgSer..... 310
|||||:|||||:|||||:|||||:|||||:|||||:
1038 CTGGCAACTCTTCGTCAATGAGGAGTCCACCATTCGCCCGAGCCCACTG 1087
311 .....GlnAsnAsnIleTyr.....SerAlaCysProArgAr 321
|||||:|||||:|||||:|||||:|||||:|||||:
1088 TGGTGAAGGGGTGGCAGGAGCTCTGTGGCCGTGCTCTGCCCTACAAC 1137
321 gAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlaProValProGlyP 338
|||||:|||||:|||||:|||||:|||||:|||||:
1138 CGTAAGGAAGAAACAAAGCATCAAGTACTGGT.....GTCTCTGGGA 1178
338 roGlyAlaProLeuProProAlaPro..... 346
|||||:|||||:|||||:|||||:|||||:|||||:
1179 AGGGCCCCAGATGGCCGCTGCCCTCTGCTGTGGACAGCGGGGTGGG 1228
347 .....LeuGlnValSerGluSerProTrp.....LeuH1 356
|||||:|||||:|||||:|||||:|||||:|||||:
1229 TTAAGGCCCACTAGAGGGCGGCTCTCCCTGTGGAGGAGCCAGGCAAC 1278
356 sAlaProSerLeuTyrThrSer 363
|||||:|||||:|||||:|||||:|||||:|||||:
1279 GGCACCTTCACTGTCTCATCTCA 1300
seq_name: N_Geneseq_36:X07407
seq_documentation_block:
ID_X07407 standard; cDNA to mRNA; 1839 BP.
AC_X07407:
DT_08-JUN-1999 (first entry)
DE_Human secretory immunoglobulin A component gene.
KW_immunoglobulin A; secretory; component; iGa; human; treatment;
KW_prevention; infection; HIV; AIDS; cold; flu; virus; gene;
KW_human immunodeficiency virus; respiratory syncytial virus; ss.
OS_Homo sapiens.
FH_Key Location/Qualifiers
FT_CDS 1..1839
FT_1/*tag= a
FT_2/product= secretory component
PN_W09857993-A1.
PD_23-DEC-1998.
PF_10-JUN-1998; U11975.
PR_19-JUN-1997; US-050969.
PA_(REGC ) UNIV CALIFORNIA.
PI_Chintalacharuvu KR, Morrison SL;
DR_WPI: 99-080950/07.
DR_P-FSDB; W95601.
PT_Producing secretory immunoglobulin in single cells - useful to
PT_produce commercial quantities of secretory immunoglobulin to prevent
PT_or treat infections
PS_Disclosure; Page 22; 39pp; English.
CC_The sequence is that of the coding region for the secretory
CC_component of human secretory immunoglobulin A (sIgA).
CC_The sequence can be used as part of a method for the
CC_production of sIg molecules. This method is useful for
CC_producing commercial quantities of sIg (especially sIgA) to treat
CC_or prevent infections. In particular, sIgA produced by the method
CC_can be used to prevent or treat infections in mammals, birds or
CC_fish; especially systemic infections or infections at a mucosal
CC_surface. It is especially useful to prevent or treat infection
CC_with human immunodeficiency virus (HIV), respiratory syncytial
CC_virus, flu virus or cold virus. The method allows production of
CC_commercial quantities of sIg molecules for therapeutic use, not
CC_previously possible; production using non-plant cells and a
CC_single cell type is more efficient than a previous multi-step
CC_process of fusing recombinant plant cells, and avoids alterations
CC_of the sIg by plant cells. sIgA molecules are more stable
CC_and resistant to proteolysis than previously used IgA molecules,
CC_and can be administered to prevent as well as to treat infections,
CC_unlike e.g. IgG and IgM molecules.
CC_Sequence 1839 BP; 423 A; 504 C; 544 G; 368 T;
SQ

```

```

alignment_scores:
  Quality: 182.00      Length: 115
  Ratio: 2.247        Gaps: 5
  Percent Similarity: 70.435  Percent Identity: 37.391

alignment_block:
  US-09-135-238B-2 x X07407 ..
  Align seg 1/1 to: X07407 from: 1 to: 1839

12  ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGlyG 28
    |||:|||||
    61  CCATATTGGT.....CCCGAG...GAGGTAATAGTGT 92
28  uLeuGlySerValThrIleLysCysProLeuPro.....G 41
    |||:|||||
    93  GGAAGTAACCTAGTGTCCATCAGTGTCTACCCACCCACCTCTGTCA 142
41  LuMetHisValArgIleLeuLysCysArgGluMetAlaGlySerGlyThr 57
    :: |||:|||||
    143  ACGGCACACCCGAGTACTGTGTCGCGCAGGAGCTAGAGTGGC... 189
58  CysGlyThrValValSerThrThrAsnPhelLysAlaGluTyrLysG 74
    |||:|||||
    190  TGCATTAACCTCACTCTCGGAGGCTAGCTCTCCAGCAATATGCGG 239
74  YArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluV 91
    |||:|||||
    240  CAGGGCTAACCTCACCACACTTCCCGAGAACGCGACATTTGTGGTGA 289
91  alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107
    :: |||:|||||
    290  TTGCCACCTGACGCGAGTACTCGCGGCGCTACAGTGTGGCTGGGC 339
108  MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
    |||:|||||
    340  ATCAATAGC...CGAGGCTGTCTTGTGATGTGTCAGCTGGAGGTC 381

seq_name: N_Geneseq_36:T30856

seq_documentation_block:
ID  T30856; standard; cDNA; 2031 BP.
AC  T30856;
DE  13-SEP-1996 (first entry)
DE  Secretory component DNA fragment 1.
KW  Secretory component; SC DNA fragment 1; crystallisation;
KW  Polymeric immunoglobulin receptor; pigR; CHO; ss.
OS  Chimeric Homo sapiens;
OS  Chimeric synthetic.
FH  Key
FT  misc_feature
FT  1..33
FT  /tag= a
FT  /note= "bases 1-33 derived from PCR with 5'
FT  primer 1"
FT  111..2015
FT  /tag= b
FT  /product= transcript from SC DNA fragment 1
FT  complement (1997..2031)
FT  /tag= c
FT  /note= "from PCR with 3' primer"
FT  WO9618734-A1.
PN  20-JUN-1996.
PD  06-DEC-1995; E04797.
PR  16-DEC-1994; EP-120019.
PA  (CIBA ) CIBA GEIGY AG.
PI  Asselbergs FAM, Gruetter M, Leist C, Van Oostrum J;
DR  WPI; 96-300651/30.
PT  prodn. of recombinant secretory protein capable of crystallisation -
PT  pref. human poly-IgR. for use in screening and binding studies
PS  Claim 5; Page 36-38; 50pp; English.
CC  SC DNA fragment 1 (T30856) was obtd. by PCR amplification of human
CC  polymeric immunoglobulin receptor (pigR) cDNA in vector pCB6 using a
CC  5' primer (T30859) hybridising to the cytomegalovirus promoter in

```

```

CC  pCB6 and a 3' primer (T30861) that generates a stop codon at
CC  position 1906 just upstream of the hydrophobic transmembrane segment
CC  of the pigR and also creates a downstream XbaI site. CHO SSF 3
CC  cells transfected with a vector carrying the amplified DNA can be
CC  used for prodn. of recombinant secretory component (SC). This can
CC  be crystallised for receptor structure studies, used to stabilise
CC  immunoglobulins, or used to screen (ant)agonists capable of
CC  modulating mucosal immune responses.
SQ  Sequence 2031 BP; 477 A; 541 C; 605 G; 408 T;

```

```

alignment_scores:
  Quality: 182.00      Length: 115
  Ratio: 2.247        Gaps: 5
  Percent Similarity: 70.435  Percent Identity: 37.391

alignment_block:
  US-09-135-238B-2 x T30856 ..
  Align seg 1/1 to: T30856 from: 1 to: 2031

12  ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGlyG 28
    |||:|||||
    171  CCATATTGGT.....CCCGAG...GAGGTAATAGTGT 202
28  uLeuGlySerValThrIleLysCysProLeuPro.....G 41
    |||:|||||
    203  GGAAGTAACCTAGTGTCCATCAGTGTCTACCCACCCACCTCTGTCA 252
41  LuMetHisValArgIleLeuLysCysArgGluMetAlaGlySerGlyThr 57
    :: |||:|||||
    253  ACGGCACACCCGAGTACTGTGTCGCGCAGGAGCTAGAGTGGC... 299
58  CysGlyThrValValSerThrThrAsnPhelLysAlaGluTyrLysG 74
    |||:|||||
    300  TGCATTAACCTCACTCTCGGAGGCTAGCTCTCCAGCAATATGCGG 349
74  YArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluV 91
    |||:|||||
    350  CAGGGCTAACCTCACCACACTTCCCGAGAACGCGACATTCGTGGTGA 399
91  alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107
    :: |||:|||||
    400  TTGCCACCTGACGCGAGTACTCGCGGCGCTACAGTGTGGCTGGGC 449
108  MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
    |||:|||||
    450  ATCAATAGC...CGAGGCTGTCTTGTGATGTGTCAGCTGGAGGTC 491

seq_name: N_Geneseq_36:T31288

```

```

seq_documentation_block:
ID  T31288; standard; cDNA; 2919 BP.
AC  T31288;
DE  24-FEB-1997 (first entry)
DE  Human poly-immunoglobulin receptor, cDNA.
KW  Human; immunoglobulin; receptor; protection protein; mutants;
KW  heavy chain; antigen binding domain; protection; pathogen;
KW  mucosal; environment; gastrointestinal; passive; immunisation;
KW  Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW  poly; sorbinus; ss.
OS  Homo sapiens.
FH  Key
FT  Location/Qualifiers
FT  235..2475
FT  /tag= a
FT  WO9621012-A1.
PN  11-JUL-1996.
PD  27-DEC-1995; U16889.
PR  30-DEC-1994; US-367395.
PR  04-MAR-1995; US-434000.
PA  (PLAN-) PLANT BIOTECHNOLOGY INC.
PA  (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA  (PLAN-) PLANET BIOTECHNOLOGY INC.

```



```

972 CTCT...AAGAGTAGAGCTTTGCAGACGCA.....GCGATTG 1006
306 roArgProArgSerGlnAsnAsnIleTyrSerAlaCysProArgAla 322
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1007 TGCATCCCGAGCTGAAGACAACATCTAC...ATTATTGAAGATAGATCT 1053
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
323 ArgGlyAlaAsp..... 326
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1054 CGAGGTGCAGATAGTCCAGAGGCTTCTGTGGGGCTTCTGCTGGG 1103
327 .....AlaAlaGlyThrGlyGluAlaProValProGlyP 338
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1104 ATTACAGAGATCGTACTGATTACAGAGTAAATACCAATCCAGCTC 1153
338 roGly.....AlaProLeuProProAlaProLeuGlnValSer 350
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1154 CTGGGAGATTTGTGTTTGTGTTCTCCAGCTGCAGTGGAGAGGTAACC 1203
351 GluSerPro 353
   :|||
1204 CTCTACCT 1212

```

seq_name: N_Geneseq_36:T96032

seq_documentation_block:

ID T96032 standard; cDNA; 2566 BP.

AC T96032;

DT 21-MAY-1998 (first entry)

DE Rat kidney injury related molecule (KIM) cDNA clone 3-2.

KW Kidney injury related molecule; KIM; rat; renal disease; injury;

KW nephritis; tissue regeneration; therapy; ss.

OS Rattus sp.

FH Key Location/Qualifiers

FT CDS 615..1538

FT CDS /*tag= a

PN WO974460-A1.

PD 27-NOV-1997.

PF 23-MAY-1997; U0930303.

PR 23-AUG-1996; US-023442.

PR 24-MAY-1996; US-018228.

PA (BIOJ) BIOGEN INC.

PI Bonventre JV, Cate RL, Hession CA, Ichimura T, Sanicola-Nadel M,

PI Wei H;

DR WPI; 98-018514/02.

DR P-PSDB; W38334.

PT DNA encoding kidney injury related molecule - which is upregulated

PT in injured or regenerating tissue, useful to promote growth of new

PT tissue and survival of damaged tissue

PS Claim 1; Page 32-34; 68pp; English.

CC cDNA clone 3-2, deposited as ATCC 98061, codes for a rat protein

CC (see W38334), designated kidney injury related molecule (KIM), that

CC is upregulated in injured or regenerating tissue. Representational

CC difference analysis was used to examine cDNA libraries prepared from

CC ischaemic and normal rat adult kidneys. 3 Cycles of hybridisation

CC and selective amplification gave 3 fragments present only in the

CC injured kidney library. Screening of the injured kidney library

CC with these fragments gave clone 3-2, as well as clone 1-7 (see

CC T96033), which is a splice variant of 3-2 encoding an identical KIM

CC polypeptide (nucleotides 136-605 of 3-2 represent an insertion),

CC and clone 4-7 (see T96034) encoding a 307-amino acid KIM protein

CC (see W38335). Clone 1-7 was used to identify a human KIM sequence

CC (see T96035). KIM nucleic acids can be used in the recombinant

CC production of KIM polypeptides in prokaryotic or eukaryotic host

CC cells. KIM, or an agonist, can be used to treat renal disease and to

CC promote the growth of new tissue or the survival of damaged tissue,

CC generally in conditions where the binding of specific ligand to KIM

CC stimulates cell growth, maintains cellular differentiation or

CC reduces apoptosis, e.g. in cases of renal failure, nephritis,

CC kidney transplants, toxic or hypoxic injury. Damage/regeneration

CC of renal cells can be determined by measuring KIM polypeptide or

CC nucleic acid, e.g. with an antisense probe, particularly to

CC diagnose or monitor the progress of disease or therapy.

CC Sequence 2566 BP; 726 A; 546 C; 591 G; 703 T;

SQ

```

alignment_scores:
  Quality: 171.00      Length: 403
  Ratio: 0.924        Gaps: 22
  Percent Similarity: 45.906  Percent Identity: 24.318
alignment_block:
US-09-135-238B-2 x T96032

```

Align seg 1/1 to: T96032 from: 1 to: 2566

```

13 ValSerGlyAlaLeuArgIleLeuPro.....GluVa 23
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
636 ATTTCAGGCCCTCTGCTCTCTTCCAGGCTCTAGATTCTTATGAAGT 685
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23 llysValGluGlyLeuGlyGlySerValThrIleLysCysProLeuP 40
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
686 A...GTGAGGGGGTGGGTGCGTACCCCTGTCACAAATCCATGTACTACT 732
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40 roGluMetHisValArgIleTyrLeuCys.....ArgGluMetAlaGly 54
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
733 CAACACGTGGAGGAATCACACGACATGTTGGGCGCGGGGCAATGCCCA 782
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 SerGlyThrCysGlyThrValValSerThrThrAsnPheIleLysAlaG 71
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
783 TATTCTAGTTGTCAAAATATATTATTGGACCAATGGATACCAAGTCAC 832
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
71 uTyrLys.....GlyArgValThrLeuLysGlnTyrProArgLysAsnL 86
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
833 CTATCGGAGCAGCGTCGATACAAACATAAAGGCGGTATTTTCAGAGGAG 882
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
86 euPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyr 102
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
883 ACGTATCTCTTGACAAATAGAGAACTCTGTTGATAGTAGTAGTGTCTGTAT 932
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 AlaCysGlyAla.....GlyMetAsnThrAspArgGlyLysThrG 116
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
933 TGTTCCCGAGTGGAGATTCCTGGATGGTTCAACGAT.....CA 970
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 nLysValThrLeuAsnValHisSerGlyTyrGluProSerTrpGluGlu 133
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
971 GAAATAGACCTTTTCATTG.....GAAGTTAAACACCAAGAAATCCCAAA 1014
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 InProMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGln 149
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1015 GTCCT....CCACACAGACCC..... 1031
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 MetProAlaTyrAlaSerSerLysPheValThrArgValThrThrPr 166
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1032 .....ACAACTACAAAGACCCCAACCAACCAAGGCCCAACTAT 1069
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166 oAlaGlnArgGlyLysValProPro.....ValHisHisSerS 179
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1070 TTCACACAGATCCACACATGTACCAACATCAACAGAGTCTCCACCTCTA 1119
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 erProThr.....ThrGlnIleThrHisArgProArgValSerArg 192
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1120 CTCCAACACCAACCAACCAACAG...ACTCACAACCAACCAACCAACT... 1163
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
193 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThrTh 209
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1164 .....ACATTTTATGCCCATGAGAC 1183
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
209 rAlaSerLysIleSerAlaLeuGluGlyLeuLysProGlnThrProS 226
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1184 AACTGCTGAGGTGACA.....GAAACTCCAT 1209
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
226 erTyr..... 227
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1210 CATATACTCTTCGACACTGGCAATGGCACTGTGACATCTCTCAGAGGAGGCC 1259
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
228 .....AsnHisHisThrArgLeuHisArgGlnArgAlaLeuAspTyrG 242

```


2

2

•
•
•
•

OM of: US-09-135-238B-2_COPY_273_390 to: Issued_Patents_NA.* out_format : pfs

Date: Sep 12, 2000 6:54 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet_p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09135238/runat_24082000_091324_17895/app_query.fasta_1.1282
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni
-GAPO=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-LOOPEXT=0.000 -GAPO=4.500 -GAPEXT=0.050 -XGAPO=10.000
-XGAPEXT=0.500 -FGAPO=6.000 -DELEXT=7.000 -START=1
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=pfs -NORM=ext -MINLEN=0 -MAXLEN=100000
-USER=US09135238 -CGN1_1_68 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-135-238B-2_COPY_273_390

Query length: 118

Database: Issued_Patents_NA.*

Database sequences: 243080

Database length: 68777915

Search time (sec): 93.130000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-824-405-5	96.50	126.18	15.46	1683	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-264-861A-10	96.00	128.90	10.92	1150	!
/cgn2_6/ptodata/1/ina/PCUTUS.COMB.seq:PCT-US95-07784-10	96.00	128.90	10.92	1150	!
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-665-259-20	95.50	124.46	19.28	1743	!
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-482-677-9	94.00	122.35	25.26	1743	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-804-227C-7	94.00	93.31	1.0e+03	44377	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-804-198-1	94.00	93.31	1.0e+03	44377	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-482-385A-5	93.00	116.93	50.66	2728	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-469-412A-8	92.00	134.42	5.38	332	!
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-07-765-830A-5	91.00	110.12	121.29	4258	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-308-949A-1	89.50	108.44	150.49	4060	!
/cgn2_6/ptodata/1/ina/backfiles1.seq:5200327-7	89.00	128.59	11.35	397	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-17	89.00	128.50	11.48	401	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-11	89.00	125.92	15.99	535	!
/cgn2_6/ptodata/1/ina/backfiles1.seq:5200327-4	89.00	125.78	16.27	543	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-15	89.00	122.00	26.44	828	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-13	89.00	121.96	26.58	832	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-11	89.00	121.89	26.80	838	!
/cgn2_6/ptodata/1/ina/backfiles1.seq:5200327-5	89.00	121.81	27.10	846	!
/cgn2_6/ptodata/1/ina/backfiles1.seq:5200327-2	89.00	121.29	28.95	896	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-7	89.00	121.25	29.10	900	!
/cgn2_6/ptodata/1/ina/backfiles1.seq:5200327-3	89.00	121.20	29.28	905	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-9	89.00	121.16	29.43	909	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-405-175A-10	89.00	116.28	55.04	1566	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-428-415-1	89.00	112.38	90.77	2419	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-379-685-1	89.00	112.38	90.77	2419	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-854-029-1	89.00	112.38	90.77	2419	!
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-09-339-964-1	88.50	105.64	215.55	4743	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-459-041A-1	88.50	99.17	494.40	9759	!
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-999-733-1	88.50	99.17	494.40	9759	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-07-642-734C-3	88.50	92.62	1.1e+03	20235	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-989-363-1	88.00	123.91	20.69	572	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-264-526-1	88.00	123.91	20.69	572	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-152-483B-2	88.00	105.83	210.38	4294	!
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-713-118-1	87.50	100.41	421.70	7266	!
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-581-148C-30	87.00	100.92	394.72	6343	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-125-468-1	87.00	86.98	2.4e+03	30001	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-474-933-1	87.00	86.98	2.4e+03	30001	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-775-009-36	86.50	105.53	218.44	3507	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-455-543A-8	86.50	99.11	497.78	7175	!

/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-193-078B-8 - 86.50 99.11 497.78 7175
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-223-305C-8 - 86.50 99.11 497.78 7175
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-149-097D-8 - 86.50 99.11 497.78 7175
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-455-543A-7 - 86.50 98.88 512.73 7362
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-193-078B-7 - 86.50 98.88 512.73 7362

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-824-405-5

seq_documentation_block:
; Sequence 5, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: dUTPase, Its Isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1683 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-824-405-5

alignment_scores:
Quality: 96.50 Length: 86
Ratio: 1.969 Gaps: 5
Percent Similarity: 56.97 Percent Identity: 33.721

alignment_block:
US-09-135-238B-2_COPY_273_390 x US-08-824-405-5/rev ..
Align seg 1/1 to reverse of: US-08-824-405-5 from: 1 to: 1683

22 ArgAlaLeuGluSerSerGlnArgProArgGlySerProArg..... 35

1020 AAAGCGCGCCCTCCCGAGGGCCCTAATGACCCACCCACCCCGCGC 971

36 .ProArgSerGlnAsnAsn.....IleTyrSerAlaCysProA 48

970 CCCCCGCGGAGAGACACCCGCTCCGCCACCTTCTCCACAGCTCAA 921

48 rgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlaPro...Val 63

[illegible]

seq_documentation_block:

; Sequence 20, Application US/08665259

; Patent No. 6028173

; GENERAL INFORMATION: CROCODILE M
 ; ADDITIONAL: I n d e x

APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy

APPLICANT: BUHN, TIMOTHY C.
APPLICANT: Connors, Timothy D.

APPLICANT: Dackowski, William R.

APPLICANT: Van Raay, Terence J.

APPLICANT: Klinger, Katherine W.

; TITLE OF INVENTION: NOVEL HUMAN COMPOSITIONS

;	TITLE OF INVENTION:	COMPOSITIONS
:	NUMBER OF SEQUENCES:	73

NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:

CORRESPONDENCE ADDRESS: ; ADDRESSEE: GENZYME CORPORATION

STREET: One Mountain Road

CITY: Framingham

69 oLeuProProAlaProLeuGlnValSerGluSerProTrpLeuHisAlap 86
|||||
1409 ATGCCACCTGCACCG.....CATAGTC 1387

86 roSerLeuLysThrSer.....CysGluTyr 94

1386 CTTCCGGCAGAACTTCTCAGGCTGATTCGGTAAC TCCACGCCG CAGGTC 1337

95 valSerLeuTyrHisGlnProAlaAla 103

1336 TGCAATGTGACTCACAGTCTCTGTGGCT 1310

name: /mnt6/ntdata/1/ine/50COMP sec.ms-09-804-2270-7

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

sequence 7, Application US/08804227C

GENERAL INFORMATION:

APPLICANT: DEHOLL, Bradley S.
APPLICANT: Kuhstoss, Stuart A.

APPLICANT: Rosteck, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.

TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:
ADDRESS: THOMAS C
PLANT 1501

STREET: LILLY CORPORATE CENTER

STATE: IN

COUNTRY: USA
ZIP: 46285

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA.

APPLICATION NUMBER: US/08/804,227C

CLASSIFICATION: 435

NAME: Plant, Thomas, G.

REGISTRATION NUMBER: 33,784
REFERENCE/DOCKET NUMBER: X-8231

TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS.

LENGTH: 44377 base pairs

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

FEATURE:	NAME/KEY:	CDS
----------	-----------	-----

LOCATION: 350..1400Z
FEATURE:

NAME/KEY: CDS
LOCATION: 14046..20036

FEATURE:	NAME/KEY:	CDS
----------	-----------	-----

LOCATION: 20110..31284
 FEATURE:

NAME/KEY:	CDS
LOCATION:	01000
	00000

FEATURE:

NAME/REF: CDS
LOCATION: 36155..41830

08-804-227C-7

gment_scores:

seq name: /cgn2 6/ptodata/1/ina/5D_COMB.seq:US-08-804-198-1


```

; MOLECULE TYPE: DNA (genomic)
US-07-765-830A-5

alignment_scores:
  Quality: 91.00      Length: 122
  Ratio: 1.596       Gaps: 5
  Percent Similarity: 46.721  Percent Identity: 26.230

alignment_block:
US-09-135-238B-2_COPY_273_390 x US-07-765-830A-5/rev ..
Align seg 1/1 to reverse of: US-07-765-830A-5 from: 1 to: 4258

      8 LysAlaLeuSerArgArgAlaArgAlaValArgMetArgAlaLe 24
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2452 CGCGCGTTGGGGTGCTTTTCAGAGAGGCGAGCCACGCTGCCCGGACTT 2403

      24 uGluSerSerGlnArgProGlySer..... 33
      | |||:|||||:|||||:|||||:|||||:|||||:|||||:
2402 GGTGTCCACGGCAGGTCCCGGAGCAGTCGCGAGCGGTGCGCCTTGAGAT 2353

      34 .....ProArgPro 36
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2352 TGGCGCCCCCGCCCCCGGAGCCTTGTGCGCCTTCTTGTGACCGCGCCC 2303

      37 .....ArgSerGlnAsnAsnIleTyrSerAlaCy 46
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2302 GCAGCGTGGGCTCGCGCAGCTCTCTCGCGCGGGGTTTCGGGACCTG 2253

      46 sProArgArgAlaArg.....GlyAlaAspAlaAlaGlyThrGly 60
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2252 TCCGAGGAAGAGCGCGCAGGTGAAGGAGACACGCGCTGCAGCAGGGG 2203

      60 luAlaProValProGlyProGlyAla..... 68
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2202 ACTGTGATGCTCCAGCCCCACCGCCCCCGCGCGGGGTGTCCTCCCA 2153

      69 ...ProLeuProAlaProLeuGlnValSerGluSerProTirLeuHi 84
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2152 AGCCAGAGCGCGCCCCACCGATGCGCGGCGACGTTGGCCCG.....CA 2109

      84 sAlaProSerLeuLys 89
      | |||:|||||:|||||:|||||:|||||:|||||:|||||:
2108 TCCACCTGCCAGCGG 2093

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-308-949A-1

seq_documentation_block:
; Sequence 1, Application US/08308949A
; Patent No. 5580703
; GENERAL INFORMATION:
; APPLICANT: Kotin, Robert M.
; APPLICANT: Berns, Kenneth I.
; APPLICANT: Linden, Ralph M.
; TITLE OF INVENTION: Human Adeno-Associated Virus Integration
; TITLE OF INVENTION: Site DNA and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/765,830A
; FILING DATE: 19910926
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 5683/92723
; TELEPHONE: 202-861-3067
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4258 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```

; MOLECULE TYPE: DNA (genomic)
US-07-765-830A-5

alignment_scores:
  Quality: 92.00      Length: 72
  Ratio: 2.875       Gaps: 3
  Percent Similarity: 44.444  Percent Identity: 37.500

alignment_block:
US-09-135-238B-2_COPY_273_390 x US-08-469-412A-8/rev ..
Align seg 1/1 to reverse of: US-08-469-412A-8 from: 1 to: 332

31 ArgGlySerProArgProArgSerGlnAsnAsnIleTyrSerAlaCysPr 47
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
331 CGCGCTCCCGCGCGCTCCG.....TGCC 306

47 o.....ArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluA 61
| |||:|||||:|||||:|||||:|||||:|||||:|||||:
305 CGTCCCGTCCCGCGCGCTCGCGCGCGCTCGCGCGCTCACCGGCTCG 256

61 laPro..... 62
| |||:|||||:|||||:|||||:|||||:|||||:|||||:
255 CCTCTAGAGCTCTCCCTCCCGCGCGCGCTCCCGGTTAATATCG 206

63 ValProGlyProGlyAlaProLeuProAlaProLeuGlnValSerGl 79
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 ACTCCGGCGCGGACGCCACGCGCGCGCGCGCGCGCGCGCGCGCGC 156

79 uSerProTirLeuHis 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 GTCGCCATGGAGAC 140

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-07-765-830A-5

seq_documentation_block:
; Sequence 5, Application US/07765830A
; Patent No. 6034231
; GENERAL INFORMATION:
; APPLICANT: TANAKA, SHOSHI
; APPLICANT: FUCHIMURA, KAYOKO
; APPLICANT: TANABAGI, YASUNORI
; APPLICANT: MATSUOTO, HISAYUKI
; APPLICANT: KANGAWA, KENJI
; APPLICANT: MINAMINO, NAOTO
; TITLE OF INVENTION: HUMAN CNP GENE AND PRECURSOR PROTEIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/765,830A
; FILING DATE: 19910926
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 5683/92723
; TELEPHONE: 202-861-3067
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4258 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/947,127
; FILING DATE: September 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Alice O.
; REGISTRATION NUMBER: 33,542
; REFERENCE/DOCKET NUMBER: ACC92-10F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-308-949A-1

alignment_scores:
  Quality: 89.50      Length: 125
  Ratio: 1.467        Gaps: 6
  Percent Similarity: 48.800  Percent Identity: 31.200

alignment_block:
US-09-135-238B-2_COPY_273_390 x US-08-308-949A-1 ..
Align seg 1/1 to: US-08-308-949A-1 from: 1 to: 4060

1 LysArgAlaValGluArgArgLysAlaLeuSerArgAlaArgArgLe 17
  :||||| :||||| :||||| :||||| :|||||
528 CGCGGAGCTGGCCGCGGAGAGCGCGCGCGCGCGCGCTTCGAGC 577
  :||||| :||||| :||||| :||||| :|||||
17 uAlaValArgMet.....ArgAlaL 24
  :||||| :||||| :||||| :||||| :|||||
578 GCGCGCGCGCTTCGCGCGCGCTGTGCGGCGCGGAGCTGGAGCGG 627
  :||||| :||||| :||||| :||||| :|||||
24 euGluSerSerGlnArgProArgGlySerProArgSerGlnAsn 40
  :||||| :||||| :||||| :||||| :|||||
628 CGTCTGATGCTGGCGCGCGCGCGCGCGCGCGCGCGCGGAGCTG 677
  :||||| :||||| :||||| :||||| :|||||
41 AsnIleTyrSerAlaCysProArgAla.....ArgG1 52
  :||||| :||||| :||||| :||||| :|||||
678 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTGAGCTCCACCA 727
  :||||| :||||| :||||| :||||| :|||||
52 yAlaAspAlaAlaGlyThrGlyGluAlaPro.....ValProGly 66
  :||||| :||||| :||||| :||||| :|||||
728 GTATCAGCGCGCTGCACAGGTGACGCGCGCGCGCGCGCGCTCC 777
  :||||| :||||| :||||| :||||| :|||||
66 roGlyAla.ProLeu.....ProProAlaProLeuGlnVa 77
  :||||| :||||| :||||| :||||| :|||||
778 CAGGTCACCTCTGCGCCACCTGGGCGATCTCTCCCGTTGCCAGT 827
  :||||| :||||| :||||| :||||| :|||||
77 lserGluSerProTrpLeuHisAlaProSerLeuLysThrSerCys 93
  :||||| :||||| :||||| :||||| :|||||
828 CTCGATCGCGCGCTGTTACTGCGCGCTGGGTTTNCACCTATGCT 877
  :||||| :||||| :||||| :||||| :|||||
94 .....TyrValSerLeuTyrHis 99
  :||||| :||||| :||||| :||||| :|||||
878 CCGCGTCCAGTCCCTTACCAT 900

seq_name: /cgn2_6/ptodata/1/ina/backfiles1.seq:5200327-7

seq_documentation_block:
; Patent No. 5200327
; APPLICANT: GARVIN, ROBERT T.; MALEK, LAWRENCE T.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR THE SECRETION OF
; BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY STIMULATING
; FACTOR (GM-CSF) AND OTHER HETEROLOGOUS PROTEINS FROM
; STREPTOMYCES
; NUMBER OF SEQUENCES: 24

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/224,568
; FILING DATE: 26-JUL-1988
; SEQ ID NO:7:
; LENGTH: 397
; 5200327-7

alignment_scores:
  Quality: 89.00      Length: 96
  Ratio: 1.745        Gaps: 7
  Percent Similarity: 53.125  Percent Identity: 37.500

alignment_block:
US-09-135-238B-2_COPY_273_390 x 5200327-7 ..
Align seg 1/1 to: 5200327-7 from: 1 to: 397

1 LysArgAlaValGluArgArgLysAlaLeuSerArgAlaArgArgLe 17
  :||||| :||||| :||||| :||||| :|||||
66 CGACGCCCATGTGACCGCTGCTGCTGCTGCGCGCGCGCGCGCGAGCT 115
  :||||| :||||| :||||| :||||| :|||||
17 u.....Ala.ValArgMetArgAlaLeuGluSerSer.GlnArgProAr 31
  :||||| :||||| :||||| :||||| :|||||
116 CGCGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG.. 163
  :||||| :||||| :||||| :||||| :|||||
31 gGlySerProArgSerGlnAsnAsnIle..... 42
  :||||| :||||| :||||| :||||| :|||||
164 .GGAGCGCTGCGCGCTGCGCGGAGACCGCGCTGCTGCTGCGGCTCAC 212
  :||||| :||||| :||||| :||||| :|||||
43 .....TyrSerAlaCysProArgArgAlaArgGlyAlaAspAlaAla 56
  :||||| :||||| :||||| :||||| :|||||
213 AGCGGTACCTCTGCGCTGCGCGGCGCGCTCGACCGCGCGGAGCTCC 262
  :||||| :||||| :||||| :||||| :|||||
57 GlyThrGlyGluAla.ProValProGly.....ProGlyAlaProLeu 72
  :||||| :||||| :||||| :||||| :|||||
263 GGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312
  :||||| :||||| :||||| :||||| :|||||
72 roAlaProLeuGlnValSerGluSerProTrp 82
  :||||| :||||| :||||| :||||| :|||||
313 CCCACCGCGCACAAAGATGTCGAAACCCCTAC 344

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-17

seq_documentation_block:
; Sequence 17, Application US/08318193
; Patent No. 5641663
; GENERAL INFORMATION:
; APPLICANT: GARVIN, ROBERT T.
; APPLICANT: MALEK, LAWRENCE T.
; TITLE OF INVENTION: AN EXPRESSION SYSTEM FOR THE SECRETION
; OF BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY
; TITLE OF INVENTION: STIMULATING FACTOR (GM-CSF) AND OTHER HETEROLOGOUS
; TITLE OF INVENTION: PROTEINS FROM STREPTOMYCES
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/318,193
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,314

```

```

; FILING DATE:
; APPLICATION NUMBER: US 07/224,568
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 18740/116 CACO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid;
; DESCRIPTION: Synthetic DNA oligonucleotide
US-08-318-193-17

alignment_scores:
    Quality: 89.00      Length: 96
    Ratio: 1.745       Gaps: 7
Percent Similarity: 53.125 Percent Identity: 37.500

alignment_block:
US-09-135-238B-2_COPY_273_390 x US-08-318-193-17 ..
Align seg 1/1 to: US-08-318-193-17 from: 1 to: 401

1 LysArgAlaValGluArgArgLysAlaLeuSerArgAlaArgLe 17
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
66 CGACGCCCATGTGACCGCTGCTGCTGCGCGCGCGCGCGAGCT 115
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

17 u.....Ala.ValArgMetArgAlaLeuGluSer.GlnArgProAr 31
| ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
116 CGCGGGGGCGGACCGCGCGCGCGCGCGCGAGTCTCGCGCGCGCGG 163
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

31 gGlySerProArgProArgSerGlnAsnAsnIle..... 42
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
164 .GGAGCGGTGCGCGCGCGCGCGCGCGAGTCTCGCGCGCGCGG 212
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

43 .....TyrSerAlaCysProArgArgAlaArgGlyAlaAspAlaAla 56
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
213 AGCGGTACCTCTCGCCTCGCGCGCGCGCGCGCTGCTGCTGCGGCTCACGG 262
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

57 GlyThrGlyGluAla.ProValProGly...ProGlyAlaProLeuProP 72
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
263 GGCACCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

72 roAlaProLeuGlnValSerGluSerProTrp 82
|| ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
313 CCCACCGCGCACAAAGAAATGTCGGAACCCCTAC 344

```

```

; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/318,193
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,314
; FILING DATE:
; APPLICATION NUMBER: US 07/224,568
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 18740/116 CACO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid;
; DESCRIPTION: Synthetic DNA oligonucleotide
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 399..533
US-08-318-193-11

alignment_scores:
    Quality: 89.00      Length: 96
    Ratio: 1.745       Gaps: 7
Percent Similarity: 53.125 Percent Identity: 37.500

alignment_block:
US-09-135-238B-2_COPY_273_390 x US-08-318-193-11 ..
Align seg 1/1 to: US-08-318-193-11 from: 1 to: 535

1 LysArgAlaValGluArgArgLysAlaLeuSerArgAlaArgLe 17
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
66 CGACGCCCATGTGACCGCTGCTGCTGCGCGCGCGCGCGAGCT 115
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

17 u.....Ala.ValArgMetArgAlaLeuGluSer.GlnArgProAr 31
| ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
116 CGCGGGGGCGGACCGCGCGCGCGCGCGCGAGTCTCGCGCGCGCGG 163
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

31 gGlySerProArgProArgSerGlnAsnAsnIle..... 42
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
164 .GGAGCGGTGCGCGCTCGCGCGGAGACCGCGCGCGCGAGTCTCGCGCGCGCGG 212
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

43 .....TyrSerAlaCysProArgArgAlaArgGlyAlaAspAlaAla 56
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
213 AGCGGTACCTCTCGCCTCGCGCGCGCGCGCTGCGACCGCGCGGACCTCC 262
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

57 GlyThrGlyGluAla.ProValProGly...ProGlyAlaProLeuProP 72
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
263 GGCACCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

72 roAlaProLeuGlnValSerGluSerProTrp 82
|| ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
313 CCCACCGCGCACAAAGAAATGTCGGAACCCCTAC 344

```

seq_name: /cgn2_6/ptodata/1/ina/backfiles1.seq:5200327-4

```
seq_documentation_block:
; Patent No. 5200327
; APPLICANT: GARVIN, ROBERT T.; MALEK, LAWRENCE T.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR THE SECRETION OF
; BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY STIMULATING
; FACTOR (GM-CSF) AND OTHER HETEROLOGOUS PROTEINS FROM
; STREPTOMYCES
; NUMBER OF SEQUENCES: 24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/224,568
; FILING DATE: 26-JUL-1988
; SEQ ID NO: 4
; LENGTH: 543
5200327-4

alignment_scores:
      Quality: 89.00      Length: 96
      Ratio: 1.745      Gaps: 7
      Percent Similarity: 53.125      Percent Identity: 37.500

alignment_block:
US-09-135-238B-2_COPY_273_390 x 5200327-4  ..
Align seg 1/1 to: 5200327-4 from: 1 to: 543

1  LysArgAlaValGluArgArgLysAlaLeuSerArgAlaArgArgLe 17
   :||||| :||||| :||||| :||| :|||||
66  CGACGCCCATGTGACCGCTGCGTGTCTGCGCGCGCGCGCGAGGCT 115
   :||||| :||||| :||||| :||||| :|||||
17  u.....Ala.ValArgMetArgAlaLeuGluSerSer.GlnArgProAr 31
   :||||| :||||| :||||| :||||| :|||||
116  CGCGCGGGCGGACCGGACCGCGCGCGAGGTCTCGCGCGGACCG.. 163
   :||||| :||||| :||||| :||||| :|||||
31  gGlySerProArgProArgSerGlnAsnAsnIle..... 42
   :||||| :||||| :||||| :||||| :|||||
164  .GGAGCGTGGCGCTCGCGCGGAGACCGCGCTGCTGCGGCTCAGG 212
   :||||| :||||| :||||| :||||| :|||||
43  .....TyrSerAlaCysProArgArgAlaArgGlyAlaAspAlaAla 56
   :||||| :||||| :||||| :||||| :|||||
213  AGCGGTACCTCTCGCGTGGCGGCGGCTCGGACCGCGCGGACCTCC 262
   :||||| :||||| :||||| :||||| :|||||
57  GlyThrGlyGluAla.ProValProGly...ProGlyAlaProLeuProP 72
   :||||| :||||| :||||| :||||| :|||||
263  GGCACGGGCGCGCGGCGGACCGCGGCGGACCGGTCGCGCGGCGCCCC 312
   :||||| :||||| :||||| :||||| :|||||
72  roAlaProLeuGlnValSerGluSerProTrp 82
   :||||| :||||| :||||| :||||| :|||||
313  CCCACCGGCACAGAATGTCCGGAACCCCTAC 344
```

OM of: US-09-135-238B-2 to: GenEmbl:* out_format : pfs
Date: Sep 12, 2000 5:34 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODE=frame+p2n.model -DEV=xlpx
-O=/cgn2_1/QPMT=fastcap -SUFFIX=p2n.rge -GAPOP=12.000
-DB=GenEmbl -QFMT=fastcap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -FGAPOP=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NRM=ext -MINLEN=0 -MAXLEN=1000000
-USER=US09135238 -CGN1_1_5120 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPPX
-WAIT -THREADS=1

Search information block:
Query: US-09-135-238B-2
Query length: 390
Database: GenEmbl:*
Database sequences: 972840
Database length: 892348106
Search time (sec): 2785.000000

score_list:	Strd Orig	zScore	EScore	Len	Documentation
Sequence	+ 2047.00	1733.42	1.9e+86	1339	! AF057557 Homo sapiens anti-Fas
gb_pr3:AF057557	+ 661.50	528.40	2.5e+21	221365	! AC023534 Homo sapiens clone
gb_htg11:AC023534	- 284.00	220.68	0.0003	58628	! AC027719 Homo sapiens chromos
gb_htg22:AC027719	+ 198.00	170.58	0.2098	3095	! U06431 Mus musculus B6/CBAF1J
gb_ro:MMU06431	+ 192.00	166.16	0.3699	2848	! AF091137 Trichosurus vulpecula
gb_cm:AF091137	+ 188.00	177.35	0.0880	417	! U83428 Mus musculus polymeric i
gb_ro:MMPGR03	+ 188.00	175.90	0.1061	505	! Y16525 Mus musculus pigR gene,
gb_ro:MMPGR03	+ 188.00	145.10	5.51	29392	! E15253 Mouse pig receptor gen
gb_pat:E15253	+ 188.00	145.10	5.51	29392	! AB001489 Mus musculus DNA for
gb_ro:AB001489	+ 187.50	165.39	0.4080	1911	! E15470 cDNA encoding immunity
gb_pat:E15470	+ 185.50	159.64	0.8532	3269	! X15741 Rat mRNA for polymeric
gb_ro:RNIGR	+ 184.00	160.53	0.7613	2461	! AB032195 Sus scrofa mRNA for p
gb_cm:AB032195	+ 183.00	156.96	1.20	3527	! L04797 Bos taurus polymeric in
gb_cm:BOVPIGR	+ 182.00	173.73	0.1400	345	! S43441 transmembrane secretory
gb_pr2:S43435503	+ 182.00	160.66	0.7483	1936	! A52094 Sequence 4 from Patent
gb_pat:A52094	+ 182.00	160.30	0.7839	2031	! A52093 Sequence 3 from Patent
gb_pat:A52093	+ 182.00	159.02	0.9239	2405	! A52091 Sequence 1 from Patent
gb_pr2:S62403	+ 182.00	157.61	1.11	2898	! S62403 transmembrane secretory
gb_pr2:S62403	+ 182.00	137.55	1.12	2919	! X73079 Homo sapiens encoding F
gb_pr1:HSP1R	+ 182.00	135.90	1.38	3630	! X81371 B.taurus mRNA for bovin
gb_cm:BPPIGR1	+ 171.00	149.35	3.19	2539	! AF035963 Rattus norvegicus kid
gb_ro:AF035963	- 167.00	112.12	378.14	221365	! AC023534 Homo sapiens clone
gb_htg11:AC023534	- 166.50	143.09	7.13	3517	! X00412 Rabbit mRNA for poly-im
gb_cm:OCIGRECR	- 159.50	118.29	171.52	42655	! AL109732 Streptomyces coelic
gb_ba1:SC7H2	+ 157.00	137.60	14.41	2526	! M24559 Human poly-Ig receptor
gb_pr1:HUMIGRPOLY	+ 152.50	146.79	4.44	456	! Y16527 mus musculus pigR gene,
gb_ro:MMPIGR5	+ 152.50	141.07	9.23	969	! AJ238323 Homo sapiens mRNA for
gb_pr2:HSAX238323	+ 152.50	141.07	9.24	970	! AJ010101 Homo sapiens mRNA for
gb_pr2:HSAX010101	+ 152.50	136.15	17.37	1857	! AJ224864 Homo sapiens mRNA for
gb_pr2:HSAX224864	+ 147.50	108.02	640.04	43599	! U15177 Human cosmid CRI-JC201
gb_pr2:HSU15177	+ 145.00	94.11	3.8e+03	207067	! AL157395 Homo sapiens chrom
gb_htg2:AL157395	+ 143.00	138.38	13.03	481	! U83430 Mus musculus polymeric i
gb_ro:MMPIGR05	- 142.50	102.95	1.2e+03	48836	! AC020151 Drosophila melanogas
gb_htg6:AC020151	- 142.50	89.56	6.8e+03	285978	! AE003551 Drosophila melanoga
gb_in2:AE003551	- 142.00	100.37	1.7e+03	64957	! AB032367 Streptomyces avermit
gb_ba1:AB032367	- 141.50	110.03	494.93	17183	! AE001986 Deinococcus radiodur
gb_ba2:AE001986	- 140.00	102.68	1.3e+03	38368	! AL031910 Leishmania major Fri
gb_in1:LMFL2743	+ 140.00	101.40	1.5e+03	45386	! AL078610 Streptomyces coelic
gb_ba1:SCH35	+ 140.00	91.53	5.3e+03	167015	! AC010208 Homo sapiens clone
gb_htg9:AC010208	- 140.00	90.07	6.4e+03	202598	! AC009813 Homo sapiens chrom
gb_htg26:AC009813	- 140.00	90.07	6.4e+03	202598	! AC009813 Homo sapiens chrom

gb_ba2:AF047717 + 139.00 113.87 302.25 7836 ! AF047717 Streptomyces chrys
gb_ba1:SC9B5 - 138.50 101.82 1.4e+03 36368 ! AL035206 Streptomyces coel
gb_ov:GDCEBP + 137.50 124.25 79.84 1686 ! X6844 G-domesticus gene fo
gb_in2:AF227923 + 137.50 120.73 125.41 2683 ! AF227923 Tribolium castaneu
gb_ba1:DVAPSAB + 137.50 118.98 156.92 3379 ! Z69372 D.vulgaris apsA and
seq_name: gb_pr3:AF057557
seq_documentation_block:
LOCUS AF057557 1339 bp mRNA PRI 04-JUN-1998
DEFINITION Homo sapiens anti-Fas-induced apoptosis (TOSO) mRNA, complete cds.
ACCESSION AF057557
VERSION AF057557.1 GI:3169292
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1339)
Hitoshi Y., Lorenz J., Kitada, S.I., Fisher, J., LaBarge, M.,
King, H.Z., Francke, U., Reed, J.C., Kinoshita, S. and Nolan, G.P.
Toso, a cell surface, specific regulator of Fas-induced apoptosis
in T cells
JOURNAL Immunity 8 (4), 461-471 (1998)
MEDLINE 98246048
REFERENCE 2 (bases 1 to 1339)
LaBarge, M. and Hitoshi, Y.
Direct Submission
AUTHORS
TITLE
Submitted (03-APR-1998) Molecular Pharmacology, Stanford
University, 300 Pasteur Dr., Palo Alto, CA 94304, USA
JOURNAL
FEATURES
Location/Qualifiers
1..1339
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q31-q32"
/cell_line="Jurkat"
/cell_type="T cell"
/note="TOSO, 2 cDNA; cloned by PCR using primers designed
from the original TOSO.1 cDNA sequence"
1..1339
/gene="TOSO"
20..1192
/gene="TOSO"
/note="transmembrane protein; Ig superfamily member"
/codon_start=1
/product="anti-Fas-induced apoptosis"
/protein_id="AAC18830.1"
/db_xref="GI:3169293"
/translation="MDFWLWPLVPLVPSGALRILPEVKVEGELGSGVTKICPLPEMHV
RIYLCEMAGSGTCGVTVSTNFIKAEYKGRVTLKQYPRKNLFLVEVTLQTESDSGVY
ACGAMNTDRGKTOKVTLNVHSEYEPSEKQPMETPKWHLPLVLFOMPAVASSKVF
TRVITPAQRGKVPVPHSSPTQTHRPVRSRASSVAGDPRFLPFTTASKISALEG
LLKPTPTSHHTLHRQALDYGSQSGREGQGHILITLGLFLALGLLVYKRAV
ERRKARRRLRAVRRALESQRPRGSRPRSONNIYSACPRRAGDAAGTGEAP
VPGGAPLPPALQVSESPWLHAPSLKTSCEYLSLYHQPAAMMEDSDSDDYINVPA"
127
/gene="TOSO"
/note="occurs in TOSO.1 cDNA"
/replace="a"
1015
/gene="TOSO"
/note="occurs in TOSO.1 cDNA"
/replace="g"
BASE COUNT 309 a 432 c 340 g 258 t
ORIGIN
alignment_scores:
Quality: 2047.00 Length: 390
Ratio: 5.262 Gaps: 0
Percent similarity: 99.744 Percent Identity: 99.744

alignment_block:

US-09-135-238B-2 x AF057557 ..

Align seg 1/1 to: AF057557 from: 1 to: 1339

1 MetAspArgTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLe 17
20 ATGGACCTTCCTGGCTTTGGCCACTTTACTTCTGCCAGTATCGGGGGCCCT 69
17 uArgIleLeuProGluValLysValGluGlyGluLeuGlyGlySerValT 34
70 GAGGATCTCTCCACAGTAAGGTAGACGGGGAGCTGGCGGATCAGTTA 119
34 hrIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 50
120 CCATCAAGTGCCTTCTCTGAAATGCATGTGAGGATATATCTGTGCGG 169
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67
170 GAGATGGCTGGATCTGGACATGTGTGTACCGTGTATCCACCACCACTT 219
67 elLeLysAlaGlyLysGlyArgValThrLeuLysGlnTyrProArgL 84
220 CATCAAGGCAGATACAAAGGGCCGAGTTTACTCTGAAGCAATACCCACGCA 269
84 ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
270 AGAATCTCTTCTAGTGGAGGTAAACACAGCTGACAGAAAGTACAGCGGA 319
101 ValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117
320 GTCTATGCTCGGAGCGGCATGAACACAGACCGGGGAAAGACCCAGAA 369
117 svAlThrLeuAsnValHisSerGluTyrGluProSerTrpGluGlnP 134
370 AGTCACCTGAAATGTCACAGTGAATACGAGCCATCATGGGAAGAGCAGC 419
134 roMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMet 150
420 CAATGCCTAGACTCCAAAATGGTTTCATCTGCCTATTGTTCCAGATG 469
151 ProAlaTyrAlaSerSerLysPheValThrArgValThrThrProAl 167
470 CTGCTATGCTGAGTCTTCCAAATTCGTAAACAGAGTTACCAACCCAGC 519
167 aGlnArgGlyLysValProProValHisHisSerSerProThrThrGlnI 184
520 TCAAAGGGGCAAGTCCCTCCAGTTCACCACTCTCTCCCCACCACCCAAA 569
184 leThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLys 200
570 TCACCCACGGCCCTCCAGTGTCCAGAGCATCTTCAGTAGCAGGTGACAAG 619
201 ProArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuGl 217
620 CCCCAGAACTTCCTGCCATCCACTACAGCTCAAAATCTCAGCTCTGGA 669
217 uGlyLeuLeuLysProGlnThrProSerTyrAsnHisHisThrArgLeuH 234
670 GGGGTGCTCTCAAGCCCGACGCCAGCTACAAACCCACCAACCCAGGCTGC 719
234 lsArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGlyGln 250
720 ACAGGCAGAGACACTGGACTATGGCTCAGACTGCGGAGGAAGGCCAA 769
251 GlyPheHisIleLeuIleProThrIleLeuGlyLeuPheLeuLeuAlaLe 267
770 GGATTCACATCTGTATCCGACCATCTCTGGGCTTTTCTCTGCTGGCACT 819
267 uLeuGlyLeuValValLysArgAlaValGluArgGlyLysAlaLeuSerA 284
820 TCTGGGCTGCTGGTGAAAGGGCGGTGAAGAGGAGGAAGCCCTCTTCCA 869

284 tgArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln 300
870 GGCGGGCCCGCCGACTGGCCGTGAGATGGCGCCCTGGAGAGCTCCCG 919
301 ArgProArgLysProArgProArgSerGlnAsnAsnIleTyrSerAl 317
920 AGGCCCGCGGGTGC CGCGACCGCTCCCAAAAACAACATCTACAGCC 969
317 acYsProArgArgAlaArgGlyAlaAspAlaGlyThrGlyGluAlaP 334
970 CTGCCCGCGCGCGCTCGTGAGCGGACGCTCGAGGCACAGGGGAAAGCC 1019
334 roValProGlyProGlyAlaProLeuProAlaProLeuGlnValSer 350
1020 CCGTTCGCCGCCCGGAGCGCGTTCGCCCGCCCGCTGCAGGTGTCT 1069
351 GluSerProTrpIleHisAlaProSerLeuLysThrSerCysGluTyrVa 367
1070 GAATCTCCCTGGCTCCATGCGCCATCTCTGAAGACCACTGTGAATACGT 1119
367 lSerLeuTyrHisGlnProAlaAlaMetMetGluAspSerAspSerAspA 384
1120 GAGCTCTACCAACAGCTCGCCCATGATGGAGGACAGTGAATCAGATG 1169
384 spTyrIleAsnValProAla 390
1170 ACTACATCAATGTTCTCTGCC 1189
seq_name: gb_hhg11:AC023534
seq_documentation_block:
LOCUS AC023534.2 221365 bp DNA HTG 02-MAR-2000
DEFINITION Homo sapiens clone RP11-462N18, *** SEQUENCING IN PROGRESS ***, 61
unordered pieces.
ACCESSION AC023534.2 GI:7143452
VERSION AC023534.2
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 221365)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome, clone RP11-462N18
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 221365)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhaltier,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Laroque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivari,T.M.,
Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
TITLE
JOURNAL
COMMENT Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2000 this sequence version replaced gi:6978210.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5244
Center clone name: 462_N_18
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1108: contig of 1108 bp in length
* 1109 1208: gap of 100 bp
* 1209 2292: contig of 1084 bp in length
* 2293 2392: gap of 100 bp
* 2393 3673: contig of 1281 bp in length
* 3674 3773: gap of 100 bp
* 3774 4947: contig of 1174 bp in length
* 4948 5047: gap of 100 bp
* 5048 5104: contig of 1057 bp in length
* 5105 6204: gap of 100 bp
* 6205 7259: contig of 1055 bp in length
* 7260 7359: gap of 100 bp
* 7360 8667: contig of 1308 bp in length
* 8668 8767: gap of 100 bp
* 8768 10566: contig of 1799 bp in length
* 10567 10666: gap of 100 bp
* 10667 12173: contig of 1507 bp in length
* 12174 12273: gap of 100 bp
* 12274 13643: contig of 1370 bp in length
* 13644 13743: gap of 100 bp
* 13744 15215: contig of 1472 bp in length
* 15216 15315: gap of 100 bp
* 15316 16603: contig of 1294 bp in length
* 16610 16709: gap of 100 bp
* 16710 18085: contig of 1376 bp in length
* 18086 18185: gap of 100 bp
* 18186 19668: contig of 1483 bp in length
* 19669 19768: gap of 100 bp
* 19769 22083: contig of 2315 bp in length
* 22084 22183: gap of 100 bp
* 22184 24052: contig of 1869 bp in length
* 24053 24152: gap of 100 bp
* 24153 25902: contig of 1750 bp in length
* 25903 26002: gap of 100 bp
* 26003 27639: contig of 1637 bp in length
* 27640 27739: gap of 100 bp
* 27740 29377: contig of 1638 bp in length
* 29378 29477: gap of 100 bp
* 29478 30910: contig of 1433 bp in length
* 30911 31010: gap of 100 bp
* 31011 31225: contig of 215 bp in length
* 31226 31325: gap of 100 bp
* 31326 32890: contig of 1565 bp in length
* 32891 32990: gap of 100 bp
* 32991 34812: contig of 1822 bp in length
* 34813 34912: gap of 100 bp
* 34913 36817: contig of 1905 bp in length
* 36818 36917: gap of 100 bp
* 36918 38575: contig of 1658 bp in length
* 38576 38675: gap of 100 bp
* 38676 41309: contig of 2634 bp in length
* 41310 41409: gap of 100 bp
* 41410 43344: contig of 1935 bp in length
* 43345 43444: gap of 100 bp
* 43445 45853: contig of 2409 bp in length
*
* 45854 45953: gap of 100 bp
* 45954 48434: contig of 2481 bp in length
* 48435 48534: gap of 100 bp
* 48535 51026: contig of 2492 bp in length
* 51027 51126: gap of 100 bp
* 51127 54369: contig of 3243 bp in length
* 54370 54469: gap of 100 bp
* 54470 56661: contig of 2192 bp in length
* 56662 56761: gap of 100 bp
* 56762 59097: contig of 2336 bp in length
* 59098 59197: gap of 100 bp
* 59198 61551: contig of 2354 bp in length
* 61552 64257: contig of 2606 bp in length
* 64258 64357: gap of 100 bp
* 64358 67511: contig of 3154 bp in length
* 67512 67611: gap of 100 bp
* 67612 70329: contig of 2718 bp in length
* 70330 70429: gap of 100 bp
* 70430 73572: contig of 3143 bp in length
* 73573 73672: gap of 100 bp
* 73673 77203: contig of 3531 bp in length
* 77204 77303: gap of 100 bp
* 77304 80599: contig of 3296 bp in length
* 80600 80699: gap of 100 bp
* 80700 84046: contig of 3347 bp in length
* 84047 84146: gap of 100 bp
* 84147 88357: contig of 4211 bp in length
* 88358 88457: gap of 100 bp
* 88458 92454: contig of 3997 bp in length
* 92455 92554: gap of 100 bp
* 92555 96892: contig of 4338 bp in length
* 96893 96992: gap of 100 bp
* 96993 101334: contig of 4342 bp in length
* 101335 101434: gap of 100 bp
* 101435 105438: contig of 4004 bp in length
* 105439 105539: gap of 100 bp
* 110414 110513: contig of 4875 bp in length
* 110514 115994: contig of 5481 bp in length
* 115995 116094: gap of 100 bp
* 116095 120447: contig of 4353 bp in length
* 120448 120547: gap of 100 bp
* 120548 124483: contig of 3936 bp in length
* 124484 124583: gap of 100 bp
* 124584 130249: contig of 5686 bp in length
* 130250 130349: gap of 100 bp
* 130350 135462: contig of 5113 bp in length
* 135463 135562: gap of 100 bp
* 135563 139847: contig of 4285 bp in length
* 139848 139947: gap of 100 bp
* 139948 146731: contig of 6784 bp in length
* 146732 146831: gap of 100 bp
* 146832 151469: contig of 4638 bp in length
* 151470 151569: gap of 100 bp
* 151570 159901: contig of 8332 bp in length
* 159902 160001: gap of 100 bp
* 160002 168308: contig of 8307 bp in length
* 168309 168408: gap of 100 bp
* 168409 176195: contig of 7787 bp in length
* 176196 176295: gap of 100 bp
* 176296 189303: contig of 13008 bp in length
* 189304 189403: gap of 100 bp
* 189404 202135: contig of 12732 bp in length
* 202136 202235: gap of 100 bp
* 202236 221365: contig of 19130 bp in length.
*
FEATURES
      Location/Qualifiers
      source
      1..221365
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="RP11-462N18"
      /clone_lib="RPC1-11 Human Male BAC"
      misc_feature
      1..1108

```

```
/note="assembly_fragment"
1209..2292
/note="assembly_fragment"
2393..3673
/note="assembly_fragment"
3774..4947
/note="assembly_fragment"
5048..6104
/note="assembly_fragment"
6205..7259
/note="assembly_fragment"
7360..8667

alignment_scores:      Quality: 661.50      Length: 508
                        Ratio: 3.410      Gaps: 8
Percent Similarity: 38.189      Percent Identity: 34.252

alignment_block:
US-09-135-238B-2 x AC023534      ..
Align seg 1/1 to: AC023534      from: 1 to: 221365

2 AspArgTrpLeuTrp..... 6
|||||
33094 GACGGCTGGAGCTGGGAAGTAGATGGGGCTGGGATATAGTTAGAGATGGG 33143
7 ....ProLeuTyrPheLeuPro.....ValSerGlyAlaLeuArgI 19
|||||:|:|
33144 GCTCCCTTCGATTATATCTCTCTCCCTGCAGTATCGGGGCCCTGAGGA 33193
19 leuLeuProGluValLysValGluGluGluLeuGlyLysSerValThrIle 35
|||||
33194 TCCTCCCAAGTAAGAGTAAAGGTAGAGGGGGAGCTGGCGGATCAGTTACCATC 33243
36 LysCysProLeuProGluMetHisValArgIleTyrLeuCysArgGluMe 52
|||||
33244 AAGTGGCCACTTCTCGAATGCATGTGAGGATATATCTGTGCGGGGAGAT 33293
52 tAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPheIleL 69
|||||
33294 GCCTGGATCTGGAAACATGGGTACCGTGGTATCCACCACCAACTTCATCA 33343
69 ysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsn 85
|||||
33344 AGGCAGATAACAAGGCCCGAGTTACTCTGAAGCAATACCCACGCAGAAAT 33393
86 LeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValty 102
|||||
33394 CTGTTCTAGTGGAGTAAACACAGCTGCAGAAAGTGACAGCGAGTCTA 33443
102 tAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValt 119
|||||
33444 TGCCTGGAGCGGGCATGAACACAGACAGACCGGGGAAAGACCCAGAAAGTCA 33493
119 hrLeuAsnValHis..... 123
|||||
33494 CCCTGAATGTGCACAGTGGTAGGTTCCTCCGCTGANTGGAGGCTCAGACT 33543
123 ..... 123
33544 ACCCAGAAAATGTTTCACTTGGGAAGATAGGAACAGACATCTAACTGTGA 33593
123 ..... 123
33594 CAGAGTCTAGACTCTTACACTGTGGCCATATAAGCCACACATAATGATGCA 33643
123 ..... 123
33644 GATCCATATGAACCTCATGATCAATTACCTCCATGTGTCACACAAATTTT 33693
123 ..... 123

33694 CTACCCAAACATGTATGAAAGACAAGATAGGTGCTCTCTTTGTATGTT 33743
123 ..... 123
33744 CATCTACCCAGGCGCATGAGGGGAGTAGAGGCTTGAGATTGAGGAGGC 33793
123 ..... 123
33794 TCTGAGATGAGGAAATTTGGAAGCAAAAGAAATAACTATCCAAACGTAAT 33843
123 ..... 123
33844 TGTCCCATGAGAAATAAAGAGAAAGACACTAATTTATCCAGCCACATTTAT 33893
123 ..... 123
33894 AGGCCAATCTTATTACTCTGTGACCTGATCCTTTATTCTCTGGAACCCC 33943
123 ..... 123
33944 GCCTTTCTCCTTCTCCTCCTGACTGAGTTCCOCACACATATGACGACGTAATT 33993
123 ..... 123
33994 CAACAGCCTAGAAATCCCAATGACCTCTCTGGTTTTCGAGGGAAATGTTGA 34043
123 ..... 123
34044 CTGGACTAGGGGAACCAAGTATTTCTTCCAGGGAAGCCCCAACCCAGCTC 34093
123 ..... 123
34094 CTTCTCTGCTTCTTTTCAGTGGACCCTGGCCCTCATCCCTCCTCTGTCATCC 34143
123 ..... 123
34144 ACTTTGGCTGAATTAGTCTTCAGTTTACACCAAGCCTCAGTCTCCCCCTA 34193
124 .....Ser..GluTyrGluProSerTr 130
:|:|:|
34194 GAGACCACCAACATATTCCTCTTCCTTCCATAGATAGATAGAGCCATCATG 34243
130 pGluGluGlnProMetProGluThrProLysTrpPheHisLeuProTyrL 147
|||||
34244 GGAGAGCAGCAATGCTCGAGACTCCAAATGGTTTCATCTGCCCCATTT 34293
147 euPheGlnMetProAlaTyrAlaSerSerSerLysPheValThrArg... 162
|||||
34294 TGTTCAGATGCCTGCATATGCCAGTTCTTCCAAATTCGTAACCAGAGGT 34343
163 ValThrThrProAlaGlnArgGlyLysValPro..... 173
|||:|:|:|
34344 CAGTTCACCATAGCAGGGTAGAGAGGTGGGTGAGATGCTGTGGCAATG 34393
174 .....ProValHisHisSerSerProThrT 182
|||||
34394 TGAATAAGGAGGAGGGGTCTATGTTCACTTCAT..... 34427
182 hrGlnIleThrHisArgProArgValSerArgAlaSerSerValAlaGly 198
:|:|:|:|
34428 ..AGAGTGGCTGGCAGAGCTGAGCTGAGTCTTAGGTGTGTATTCCTCGCC 34475
199 AspLysProArgThrPheLeuProSerThrThrAlaSerLysIleSerAl 215
:|:|:|:|
34476 TTATAGAGAGGCACATACCTCCCTGAGATCTGGGTCATATCTCTCAAG 34525
215 aLeuGluGlyLeuLeuLysProGlnThrProSerTyrAsnHisHis..... 230
|||:|:|:|
34526 CCTGTCCGACAGTCAAGTCCCTGGTCCGTAGCTGCAGAACTTACATATTA 34575
231 ..ThrArgLeuHisArgGlnArg 237
|||||
34576 CCACCAGGTGGCACCACAAGCAA 34598
```



```
seq_name: gb_htg22:AC027719
seq_documentation_block:
LOCUS AC027719 58628 bp DNA HTG 01-APR-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-258P9 map 8, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC027719
VERSION AC027719.1 GI:7382531
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 58628)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-258P9
Unpublished
2 (bases 1 to 58628)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferrera,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Melgrim,J., Menelus,L., Mihova,P., Miranda,C., Menga,Y., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,J., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7449
Center clone name: 258_P_9
-----
* NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 671: contig of 671 bp in length
* 672 771: gap of 100 bp
* 772 1469: contig of 698 bp in length
* 1470 1569: gap of 100 bp
* 1570 2261: contig of 692 bp in length
* 2262 2361: gap of 100 bp
* 2362 3051: contig of 690 bp in length
* 3052 3151: gap of 100 bp
* 3152 3916: contig of 665 bp in length
* 3917 4597: contig of 681 bp in length
* 4598 4697: gap of 100 bp
* 4698 5395: contig of 698 bp in length
* 5396 5495: gap of 100 bp
* 5496 6185: contig of 690 bp in length
* 6186 6285: gap of 100 bp
* 6286 6969: contig of 684 bp in length
* 6970 7069: gap of 100 bp
* 7070 7780: contig of 711 bp in length
* 7781 8590: contig of 710 bp in length
* 8591 8690: gap of 100 bp
* 8691 9372: contig of 682 bp in length
* 9373 9472: gap of 100 bp
* 9473 10162: contig of 690 bp in length
* 10163 10262: gap of 100 bp
* 10263 10944: contig of 682 bp in length
* 10945 11044: gap of 100 bp
* 11045 11747: contig of 703 bp in length
* 11748 11847: gap of 100 bp
* 11848 12547: contig of 700 bp in length
* 12548 12647: gap of 100 bp
* 12648 13336: contig of 689 bp in length
* 13337 13436: gap of 100 bp
* 13437 14103: contig of 667 bp in length
* 14104 14203: gap of 100 bp
* 14204 14872: contig of 669 bp in length
* 14873 14972: gap of 100 bp
* 14973 15682: contig of 710 bp in length
* 15683 15782: gap of 100 bp
* 15783 16491: contig of 709 bp in length
* 16492 16591: gap of 100 bp
* 16592 17290: contig of 699 bp in length
* 17291 17390: gap of 100 bp
* 17391 18088: contig of 698 bp in length
* 18089 18188: gap of 100 bp
* 18189 18883: contig of 695 bp in length
* 18884 18983: gap of 100 bp
* 18984 19671: contig of 688 bp in length
* 19672 19771: gap of 100 bp
* 19772 20413: contig of 642 bp in length
* 20414 20513: gap of 100 bp
* 20514 21196: contig of 683 bp in length
* 21197 21296: gap of 100 bp
* 21297 21992: contig of 696 bp in length
* 21993 22092: gap of 100 bp
* 22093 22804: contig of 712 bp in length
* 22805 22904: gap of 100 bp
* 22905 23604: contig of 700 bp in length
* 23605 23704: gap of 100 bp
* 23705 24402: contig of 698 bp in length
* 24403 24502: gap of 100 bp
* 24503 25191: contig of 689 bp in length
* 25192 25291: gap of 100 bp
* 25292 25974: contig of 683 bp in length
* 25975 26074: gap of 100 bp
* 26075 26769: contig of 695 bp in length
* 26770 26869: gap of 100 bp
* 26870 27572: contig of 703 bp in length
* 27573 27672: gap of 100 bp
* 27673 28373: contig of 701 bp in length
* 28374 28473: gap of 100 bp
* 28474 29165: contig of 692 bp in length
* 29166 29265: gap of 100 bp
* 29266 29956: contig of 691 bp in length
* 29957 30056: gap of 100 bp
* 30057 30741: contig of 685 bp in length
* 30742 30841: gap of 100 bp
* 30842 31519: contig of 678 bp in length
* 31520 31619: gap of 100 bp
```

31620 32323: contig of 704 bp in length
32324 32423: gap of 100 bp
32424 33110: contig of 687 bp in length
33111 33210: gap of 100 bp
33211 33902: contig of 692 bp in length
33903 34002: gap of 100 bp
34003 34708: contig of 706 bp in length
34709 34808: gap of 100 bp
34809 35500: contig of 692 bp in length
35501 35600: gap of 100 bp
35601 36300: contig of 700 bp in length
36301 36400: gap of 100 bp
36401 37097: contig of 697 bp in length
37098 37197: gap of 100 bp
37198 37900: contig of 703 bp in length
37901 38000: gap of 100 bp
38001 38695: contig of 695 bp in length
38696 38795: gap of 100 bp
38796 39487: contig of 692 bp in length
39488 39587: gap of 100 bp
39588 40291: contig of 704 bp in length
40292 40391: gap of 100 bp
40392 41101: contig of 710 bp in length
41102 41201: gap of 100 bp
41202 41886: contig of 685 bp in length
41887 41986: gap of 100 bp
41987 42686: contig of 700 bp in length
42687 42786: gap of 100 bp
42787 43484: contig of 698 bp in length
43485 43584: gap of 100 bp
43585 44276: contig of 692 bp in length
44277 44376: gap of 100 bp
44377 45078: contig of 702 bp in length
45079 45178: gap of 100 bp
45179 45869: contig of 691 bp in length
45870 45969: gap of 100 bp
45970 46664: contig of 695 bp in length
46665 46764: gap of 100 bp
46765 47456: contig of 692 bp in length
47457 47556: gap of 100 bp
47557 48249: contig of 693 bp in length
48250 48349: gap of 100 bp
48350 49042: contig of 693 bp in length
49043 49142: gap of 100 bp
49143 49843: contig of 701 bp in length
49844 49943: gap of 100 bp
49944 50640: contig of 697 bp in length
50641 50740: gap of 100 bp
50741 51438: contig of 698 bp in length
51439 51538: gap of 100 bp
51539 52231: contig of 693 bp in length
52232 52331: gap of 100 bp
52332 53020: contig of 689 bp in length
53021 53120: gap of 100 bp
53121 53857: contig of 737 bp in length
53858 53957: gap of 100 bp
53958 54652: contig of 695 bp in length
54653 54752: gap of 100 bp

alignment_scores:

Quality: 284.00 Length: 80
Ratio: 4.733 Gaps: 2
Percent Similarity: 75.000 Percent Identity: 73.750

alignment_block:

US-09-135-238B-2 x AC027719/rev ..

Align seg 1/1 to reverse of: AC027719 from: 1 to: 58628

2 AspArgTrpLeuTrp..... 6

||| ||| |||

40676 GACGGCTGGAGCTGGAGTAGATGGGCTGGGATAGATAGATAGATGGG 40627

7ProLeuTyrPheLeuPro.....ValSerGlyAlaLeuArgI 19
|||||: ||| |||||||||
40626 GCTCCCTCTGCTAAATCTCTCCTCTGCTAGTATCAGGGCCCTGAGGA 40577

19 leLeuProGluVallysValGluGlyGluLeuGlyGlySerValThrIle 35
||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
40576 TCTCCCAAGAAGTAAGGTAGAGGGGAGCTGGCGGATCAGTACCATC 40527

36 LysCysProLeuProGluMetHisValArgIleTyrLeuCysArgGluMe 52
||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
40526 AAGTGCCTCTCTGAAATGCATGTGAGGATATATCTGTGCGGGAGAT 40477

52 talaglySerGlyThrCysGlyThrValValSerThrThr 65
||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
40476 GCGTGGATCTGGAACATGTGTACCGTGTATCCACCACC 40437

seq_name: gb_ro:MMU06431

seq_documentation_block:
LOCUS MMU06431 3095 bp mRNA ROD 26-MAY-1995
DEFINITION Mus musculus B6/CBAF1J polymeric immunoglobulin receptor mRNA,
complete cds.
ACCESSION U06431
VERSION U06431.1 GI:458421
KEYWORDS
SOURCE mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Piskurich,J.F., Blanchard,M.H., Youngman,K.R., France,J.A. and Kaetzel,C.S.
TITLE Molecular cloning of the mouse polymeric Ig receptor. Functional regions of the molecule are conserved among five mammalian species
J. Immunol. 154 (4), 1735-1747 (1995)
MEDLINE 95138517
REFERENCE 2 (sites)
AUTHORS Piskurich,J.F., Hsieh,M., Youngman,K.R. and Kaetzel,C.S.
TITLE Molecular cloning of mouse polymeric Ig receptor
J. Immunol. 150, 38A-38A (1993)
REFERENCE 3 (bases 1 to 3095)
AUTHORS Piskurich,J.F.
TITLE Molecular Cloning and Regulation of the Polymeric Immunoglobulin Receptor
Thesis (1994) Pathology, Case Western Reserve University
REFERENCE 4 (bases 1 to 3095)
AUTHORS Piskurich,J.F.
TITLE Direct Submission
Submitted (08-FEB-1994) Janet F. Piskurich, Case Western Reserve University, Pathology, Biomedical Research Building, Cleveland, OH 44060, USA

FEATURES
source Location/Qualifiers
1..3095
/organism="Mus musculus"
/strain="B6/CBAF1J (C57 Black 6 X CBA)"
/db_xref="taxon:10090"
/clone="G20"
/clone_lib="catalog #935302, Stratagene Cloning Systems, La Jolla, CA"
/sex="female"
/tissue_type="liver"
<1..84
85..138
85..2400
/codon_start=1
/function="binds and transports polymeric immunoglobulin"
/product="polymeric immunoglobulin receptor"
/protein_id="AA067440.1"
/db_xref="GI:458422"
/translation="MRLLFLVTFSGVSTKSPIFGQEVSSIEGDSVSTCYYPD
TSVNHTRKTKWQKQASGMCTLLISSNGYLSKESGRANLIFNPNFTINIEQLTQ
DDTSYKGLGTSFVSLVSPQLPSDTHVTYKIDGRNVTIECPFKRNVP
SKSLCKKTNQSCVELIDSTEKVNPSYIGRAKLFMKGTDLTVFYVNIHLTHNDAGLY

ICQAGEPSADKKNVLDQLVLAPELlyKDLRSVTPCEDLGREVANEAKYLCLRMNKE
TCDVILINTLGRDPDFEGRILLITPKDDNGRFSVLITGLRKEDAGHYOCGAHSSGLPQE
GWPIQTQWLFNEESTIPNRSVVKVGTGSAVATACYPNKESSSLKYCWRESDGNG
HCPALVGTQAOVQREYEGRLALEDPQNGNYTILNLTEDAGFTWCLTNGDSRRT
TLEQVAETREPNEVTPONATVLTGETVTSYQEKYQKWKWSNGCHIL
PSHDEGARQSSVCDQSOLVSMTLNPNVSKEDGWCYQKQGTGETTAYIAVEE
TRGSHVNPPTDANARAKVLEEVDSSISEKENKAIPNPGPFANERETQNYRDOAQ
ENRASGDAGSADGQSSSSKVLFTPLVLGLAVGAIAVWVARVHRKKNVDRMSIS
SYRTDISMADFKNSRDLGGNDMGASPDQOTVTEGKDEIVTITECTABPEESKKAKR
SSKEEDMAYSAFLQSSSTAAQVHDGPQA"
139..2397
139..2019

mat_peptide
misc_feature

/note="ectoplasmic region, binds polymeric immunoglobulin"
2020..2088

/note="transmembrane segment of the receptor"

misc_feature

/note="cytoplasmic tail of the receptor"

2401..3095

3'UTR 861 a 796 c 784 g 654 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 198.00 Length: 168

Ratio: 1.980 Gaps: 5

Percent Similarity: 59.524 Percent Identity: 31.548

alignment_block:

US-09-135-238B-2 x MMU06431 ..

Align seg 1/1 to: MMU06431 from: 1 to: 3095

10 PheLeuProValSerGlyAlaLeuArgIleLeuProGluValLysValG1 26

121 TTTTCAGGGGCTCCACAAAACCCCATTTGGTCCCCAGAGGTGAG 170

26 uGlyGluLeuGlyGlySerValThrIleLysCysProLeuProGlu... 41

171 TAGTATAGAGCGGACTCTGTTTCCATCAGCTGCTACTACCCAGACACCT 220

42MetHisValArgIleTyrLeuCysArgGluMetAlaGlySer 55

221 CTGTCAACCGGCACACCCGGAATACTGTTGCCGACAAAGGAGCC...AGC 267

56 GlyThrCysGlyThrValValSerThrThrAspPheIleLysAlaGluTyr 72

268 GGCATGTGCACACGCTCATCTCTTCAATGGCTACCTCTCCAAGGAGTA 317

72 rLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuV 89

318 TTCAGGCAGAGCAACCTCATCACTTCCAGAGACACACATTTGTGA 367

89 alGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGly 105

368 TTAACATTGAGCAGCTCACCCAGGAGGACACTGGGAGCTACAAGTGTGC 417

106 AlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVa 122

418 CTGGGT...ACCAAGTAAACGAGCGCTGCTCTTCGATGTCAGCTGGAGGT 464

122 lHisSerGluTyrGluProSerTyrGluGluGlnProMetProGluThrP 139

465 C.....AGCCAGGTTCCTGAGTTGC 484

139 roLysTrpPheHisLeuProTyrLeuPheGlnMetProLalaTyrAlaSer 155

485 CGAGTGACACCCAGCTC.....TACACAAAG 510

156 SerSerLysPheValThrArgValThrThrProAlaGlnArgGlyLysVa 172

511 GACATAGGCAGAAAATGTGACCATTTGAATGCCCTTTTCAAAAGGGAGAATGT 560

172 lPro 173

||||
561 TCCC 564

seq_name: gb_om:AF091137

seq_documentation_block:

LOCUS AF091137 2848 bp mRNA MAM 01-JUL-1999

DEFINITION Trichosurus vulpecula polymeric-immunoglobulin receptor precursor

(PIGR) mRNA, complete cds.

ACCESSION AF091137

VERSION AF091137.1 GI:5305512

SOURCE Trichosurus vulpecula.

ORGANISM Trichosurus vulpecula

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Metatheria; Diprotodontia; Phalangeridae; Trichosurus.

1 (bases 1 to 2848)

Adamski,F.M. and Demmer,J.

Cloning and characterization of pigr and J chain of the marsupial,

Trichosurus vulpecula (brush-tailed possum)

Unpublished

2 (bases 1 to 2848)

Adamski,F.M. and Demmer,J.

Direct Submission

Submitted (09-SEP-1998) Dairy Science, AgResearch Ruakura, East

Street, Hamilton, New Zealand

Location/Qualifiers

1..2848

/organism="Trichosurus vulpecula"

/db_xref="taxon:9337"

/tissue_type="mammary gland"

1..2848

/gene="PIGR"

149..2350

/gene="PIGR"

/codon_start=1

/product="polymeric-immunoglobulin receptor precursor"

/protein_id="AAD41688.1"

/db_xref="GI:5305513"

/translation="MAFLACLALLPVVSMKSPIFGPKQVTVGGSVSTQCFYPSS

SVNRHGRKTYFCLNQRCSETIVSSNGFVSERFSRAKLTFNPGNNSLILQISLEKE

DILYKCGLGTSNRGLSDITILEVGEPLNPNTVEITVEGKTVSINCPFOEQNTQD

RKFLCKKDGESCALVIDSQVQPDYTGARLSISGTSRFEVTVISQIKRQDGMVY

CGVGEDSDTGIQKNVDLKLLEPEPELLYAGLGSVTLNLCALGSTVASVPKFCQMRK

ETCDLVINSKGFNTNATHGRILFSTHTGSPSIMTOVRKEDEGVYHCGAEDGQPS

EEGPTRALQLFVSEETVPKPLVYKVGSGSVTITCYDPKKNTLYWKWEGSSH

CTKLVDLSUGMDESEYGRVALDEPENGIITVILNLTQPDAGIYWCLSNGEHNKSS

VRIENDGQPLLIAPKVTQAQLGSLTISCHEPCKFYSEYKWCWSNQGCETLPTQE

EGSSQAFVDCNQNSRNVSILTNSVTRDHEGWYCWGKNGQNYGETIAVSVSVEEVSG

NAIOPTNAVNLNEDAVEPKVRKEIEVPTDLGSTEBHSGSSVLTPLVALVLTGVA

VALGIIKABRWRFSDRVSGSYRTDLSMSELENNPRQFGANEMDASVQETILGGEDE

IITOTENPEEIOQPKQVRKSSREEDMATAFLLOSNNMANEIPLODPSNA"

BASE COUNT 769 a 690 c 722 g 667 t

ORIGIN

alignment_scores:

Quality: 192.00 Length: 204

Ratio: 1.684 Gaps: 8

Percent Similarity: 55.882 Percent Identity: 30.392

alignment_block:

US-09-135-238B-2 x AF091137 ..

Align seg 1/1 to: AF091137 from: 1 to: 2848

17 LeuArgIleLeuProGluVal..... 23

173 CTGTGCTCTCCTCCCTCTGTTCTTATGAGAGTCCCATATTTGCCCCAA 222

24 .LysValGluGlyGluLeuGlyGlySerValThrIleLysCysProLeup 40

223 GCAAGTGACAGGAGTGGGAAGGAGGATCTGTCTCCATCCAGTCTTCTACC 272


```

/chrosome="1"
/map="le3"
/sex="female"
/tissue_type="liver"
2470..2474
2580..2585
2878..2882
3176..3180
3251..3255
3350..3355
3389..3501
/note="based on the putative transcription start"
3389..3501
/number=1
3502..11067
/number=1
11124..27591
/genes="pigr"
Join(11124..11166,18192..18536,21130..21783,23022..23354,
23908..24240,25133..25325,25686..25813,26183..26314,
26426..26484,27496..27591)
/genes="pigr"
/functions="mediates secretion of IgA and IgM"
/codon_start=1
/product="polyimmunoglobulin receptor"
/protein_id="BAA24431.1"
/db_xref="GI:2804246"
/translation="MRLXLTLLVTFVSGVSTKSPFGPQEVSSIEGDSVITCYYPD
TSVNRHTRKYKMGQAGMCTLLSSNGYLSKEYSGRANLNPNENNVINIQLTQ
DDTSYKCGIGTSNRGLSFDVLSQVPELSDHYTRDGRNVTIECPFKRENAP
SKSLCKTNQSCELVIDSTKYNPSYIGRAKLEFMKGDULFVFNISHLTHNDAGLY
ICQAGEPSADKKNLDQLYLAPEPELLDKLRSVTFECDDIGREANEAKYLCRNKE
TCDVLTQLEKRDPEGRILITPKDDNGRFVLIJTLRKEDAGHYQCGAHSSGLPQE
GWPIQVQWLEFNEESTIPNRSVYKGTGGSVAIACPYNPKSSSLKYWCWREGDNG
HCPVLVGTQAOVEEYEGRLALFDQNGTFTVILNQLTTEDAGFYWCLTNGDSRWT
TIELQVAETREPNEVTPQNAVLGETFTVSCYPCKFVSOEYKWKWSKKSCHIL
PSHDGAROSSVSCDQSSOLVMTLPVSKDEGWKGVKQGTGETATYIAYVEE
RTRGSHVNPTDANAKVALEEFVDSISSEKENKAIPNPGPFANEREIQNVQDQQAQ
ENRAGDAGSADGQSSSKVLESLYPLGLVLAGAIAVWVRVRRHKNVDRMSIS
SYRTDISMDFKNSRDLGGNDGASPDYQQIVIEGKDEIVTTTECTAEPEESKKAR
SKSEADMAYSAPFLQSSSTIAAQVHDGPQEA"
11124..11166
/genes="pigr"
/number=2
11167..18191
/genes="pigr"
/number=2
18192..18536
/genes="pigr"
/number=3
18537..21129
/genes="pigr"
/number=3
21130..21783
/genes="pigr"
/number=4
21784..23021
/genes="pigr"
/number=4
23022..23354
/genes="pigr"
/number=5
23355..23907
/genes="pigr"
/number=5
23908..24240
/genes="pigr"
/number=6
24241..25132
/genes="pigr"
/number=6
25133..25325
/genes="pigr"

```

```

/number=7
25326..25685
/genes="pigr"
/number=7
25686..25813
/genes="pigr"
/number=8
25814..26182
/genes="pigr"
/number=8
26183..26314
/genes="pigr"
/number=9
26315..26425
/genes="pigr"
/number=9
26426..26484
/genes="pigr"
/number=10
26485..27495
/genes="pigr"
/number=10
27496..29008
/number=11
27592..29008
/notes="based on the polyadenylation site"
polyA_signal 28983..28988
polyA_site 29008
BASE COUNT 8318 a 6747 c 6514 g 7813 t
ORIGIN

alignment_scores:
    Quality: 188.00      Length: 97
    Ratio: 2.648        Gaps: 3
    Percent Similarity: 73.196    Percent Identity: 40.206

alignment_block:
US-09-135-238B-2 x AB001489 ..

Align seg 1/1 to: AB001489 from: 1 to: 29392

30 GlyGlySerValThrIleLysCysProLeuProglu.....Me 42
||| |||||:||||| ||| |||:
18245 GGGACTCTGTTTCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 18294
42 tHisValArgIleTyrLeuCysArgIleMetAlaGlySerGlyThrCysG 59
|||:||||| ||| |||: ||| |||: |||
18295 GCACACCCGGAATACTGTGCGCAAGAGGCC...AGCGCATGTGCA 18341
59 lYThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 75
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18342 CAACGCTCATCTCTCAATAGGCTCTCTCCAAGGAGTATTTCAGGCAGA 18391
76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
18392 GCCAACTCATCAACTTCCAGAGAACACACATTTGTGATTAACTTGA 18441
92 rGlnLeuThrGluSerSerGlyValTyrAlaCysGlyAlaGlyMeta 109
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18442 GCAGCTCACCCAGCAGACACTGGGAGCTCAAGTGTGGCTGGGT...A 18488
109 snThrAspArgGlyLysThrGlnLysValThrLysValThrLeuAsnVal 122
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18489 CCAGTAACCGAGCGCTGTCTTCGATGTCAGCTGGAGGTC 18529

seq_name: gb_pat:E15470

seq documentation_block:
LOCUS E15470
DEFINITION cDNA encoding immunity related factor.
ACCESSION E15470
VERSION E15470.1 GI:5710153

```

```

KEYWORDS JP 1998072495-A/1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1911)
AUTHORS Miyabayashi.T. and Sakano.S.
TITLE IMMUNITY-RELATED FACTOR
JOURNAL Patent: JP 1998072495-A 17-MAR-1998;
ASAHII CHEM IND CO LTD
COMMENT OS Homo sapiens (human)
PN JP 1998072495-A/1
PD 17-MAR-1998
PF 11-JUN-1997 JP 1997153218
PR 13-JUN-1996 JP 96P 152362
PI MIYABAYASHI TOMOYUKI, SAKANO SEIJI
PC C07K14/47,C07H21/04,C07K16/18,C12N5/10,C12N15/09, PC
C12P21/02//A61K38/00,
PC C12P21/08, (C12N5/10,C12R1:91),(C12P21/02,C12R1:91),(C12P21/08,
PC C12R1:91);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH Key Location/Qualifiers
FH 1..1911
FT source /organism='Homo sapiens'
FT /tissue_type='lymph node'
FT 5'UTR 1..97
FT sig_peptide 98..145
FT /product='signal peptide of immunity related
FT protein'
FT mat_peptide 146..1693
FT /product='immunity related protein' FT CDS
FT 98..1696 /product='immunity related protein' FT 3'UTR
1697..1911.
FEATURES
source Location/Qualifiers
1..1911
/db_xref='taxon:9606'
BASE COUNT 490 a 541 c 525 g 355 t
ORIGIN

alignment_scores:
Quality: 187.50 Length: 575
Ratio: 0.906 Gaps: 20
Percent Similarity: 36.000 Percent Identity: 18.957

alignment_block:
US-09-135-238B-2 x E15470 ..
Align seg 1/1 to: E15470 from: 1 to: 1911
3 ArgTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLeuArg11 19
|||||
170 AGATGGCTGTGGAG...GGCTCTCTCCCTCCAGGACCCATCTCCGGGC 216
19 LeuProGluValLys..... 24
:::
217 CATGGGAACATCAGGCCTCTCTCGCCCTCTGCTGGGGGAGGAGCT 266
25 .....ValGluGlyGlu 28
|||||
267 CCTTTGCAGCTCCAAATTCATTGAAGGGCTCAAGGCTGGTGTGTCAGGGGAG 316
29 LeuGlyGlySerValThrIleLysCysProLeuPro.....G1 41
|||||
317 CTGGAGGAGCTGTCCACCATCAGTGCCATTATGCCCTCATCTGTCTCAA 366
41 uMethHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrC 58

```

```

: ||| ||| ||| |||||
367 CAGGCACAGAGAGTACTGTGCGCTGGGCCCCCAAGATGATCT 416
58 ysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGly 74
|| |||||
417 GCAGACCATTTGTCTCCACCAACCATGATATCACTACCATCGCTATCGTGAC 466
75 ArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluVa 91
|||||
467 CGTGTGGCCCTCAGAGACTTCCACAGAGAGCTTGTGTGTGGTGGGCT 516
91 IThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyM 108
:::
517 GTCCCACTGTCCCGGATGACATCGGATGCTACCTCTGCGGCATGGAA 566
108 etAsnThrAspArg..... 112
:::
567 GTGAAACAACATGCTGTTCTTAAGCATGAATCTGACCATCTCTGCGAGT 616
112 ..... 112
617 CCGGCCAGACCTCCCCACAGCCACTCCAGCTGCTGGGAGGCTCACCAT 666
113 .....GlyL 114
667 GAGATCCTATGAAACAGCGCTCTCCAGTGGCCAAACAGATGGACCCAGGAA 716
114 ysThrGln..... 116
717 CCACCCAGACCTTAGGACAGGGGACAGCATGGACACAGTGTCTTCCACT 766
117 .....LysValThrLeuAsnValHisSerGluTyrGluPro... 128
|||||
767 CCAGGAACAGCAAGACTACAGCTTCAGCTGAGGAAGACGAACCCAGG 816
129 .....SerTirpGluGlu..... 132
817 AGCAACACGAGGCGAGCTCCAGGACAGGCGAGCTGGGAGGGTCTTG 866
133 .....GlnProMetProGluThrProLysTirpPheHisLeuPro 145
|||||
867 TCAAGACACCTGCTCCGATTCAGAGAGTCCA..... 898
146 TyrLeuPheGlnMetProAlaTyrAlaSerSerLysPheValThrAr 162
899 .....CCTTCAAAGACAGGAAGCATGTCCAA 924
162 gValThrThrProAlaGlnArgGlyLysValProProValHisHisSerS 179
:::
925 TACAACAGAGGTGTTGGAGGGC..... 949
179 erProThrThrGlnIleThrHisArgProArgValSerArg..... 192
:::
950 ..ACCAGAAAGCTCGGTGACAAACAGGGCTAGAGCCAGCAAGGACAGGAG 997
193 ..AlaSerSerValAlaGlyAspLysProArg..... 202
:::
998 GAGATGACAACATACCAAGGCTGATAGGCCAAGGAGGACATAGAGGGGT 1047
202 ..... 202
1048 CAGGATAGCTCTTGATGAGCCAAAAGGTCTAGAACCATTTGGGCCAC 1097
203 .....ThrPheLeuProSerThrThr 209
1098 CAGCTCTGGTCTCAGAAACTTTGGCGCTGGGAAATCTCTCCCAACAAGCAAG 1147
210 AlaSerLysIleSerAlaLeuGluGlyLeuLysProGlnThrProSe 226
1148 CCAGTTTCTTAAGCAACAATCTCAGGGTTCCATTGGAGAAACAACCTCCAGC 1197
226 rTyrAsnHisHisThrArg..... 232
:::

```



```

142 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 159
      |||:::
484 CCATGTC.....TACACAAAGG 500

159 heValThrArgValThrProAlaGlnArgGlyLysValProProVal 175
      :: |||:::|||| ||| ||| :::
501 ACATAGGCAGCAACTGTGACCATCGAATGCCGTTTCAAGAGGGGAATGCT 550

176 HisHisSer_SerProThrThrGlnIleThrHisArgPro 188
      ||| ::: |||:::||||: :::|||||
551 CATAGCAAGAAATCCCTGTGTGAAGAGAGAGAGAGAGGCT 590

```

seq_name: qb_om:AB032195

<hr/>					
seq_documentation_block:					
LOCUS	AB032195	2461 bp	mRNA	MAM	03-FEB-2000
DEFINITION	Sus scrofa mRNA for poly-Ig receptor, complete cds.				
ACCESSION	AB032195				
VERSION	AB032195.2	GI:6863079			
KEYWORDS	poly-Ig receptor.				
SOURCE	Sus scrofa female cDNA to mRNA, clone:SC1.				
ORGANISM	Sus scrofa				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Cetartiodactyla; Suina; Suidae; Sus.				
REFERENCE	1 (bases 1 to 2461)				
AUTHORS	Sone,T. and Kumura,H.				
TITLE	Porcine mammary gland cDNA clone, similar to poly-Ig receptor				
JOURNAL	Published Only in DataBase (1999); In press				
REFERENCE	2 (bases 1 to 2461)				
AUTHORS	Sone,T. and Kumura,H.				
TITLE	Direct Submission				
JOURNAL	Submitted (07-SEP-1999) to the DDBJ/EMBL/GenBank databases. Teruo				
	Sone, Hokkaido University, Laboratory of Applied Microbiology,				
	Faculty of Agriculture, Kita-9, Nishi-9, Kita-ku, Sapporo, Hokkaido,				
	060-8589, Japan (E-mail:sonet@chem.agr.hokudai.ac.jp),				
	Tel:81-11-706-2502, Fax:81-11-706-4961)				
COMMENT	On Feb 3, 2000 this sequence version replaced gi:5881588.				
	Sequence updated (01-Feb-2000).				
FEATURES	Location/Qualifiers				
<hr/>					

BASE COUNT	608 a	699 c	708 g	446 t
ORIGIN				

alignment_scores:		
Quality:	184.00	Length: 120
Ratio:	2.217	Gaps: 4

```

Percent Similarity: 69.167      Percent Identity: 34.167

alignment_block:
US-09-135-238B-2 x AB032195      ..

Align seg 1/1 to: AB032195 from: 1 to: 2461

17  LeuArgIleLeuProGluValLysValGluGlyGluLeu..... 29
85  CTGGCTATCTTCCCATCGAGTGGTCTCCATGAAGAGTCCCATATCGGTCCCA 134
30  .....GlyGlySerValThrIleLysCysProLeuP 40
135 GGATGTGAGCAGCGTGAAGCAGCTCGGTGTCATCATGATGCTACTACC 184
40  ro.....GluMetHisValArgIleYrLeuCysArgGluMet 52
185 CAGCCACCTCCGTCACACGGCATCTCTCGGAAGTACTGGTGCAGA...ATA 231
53  AlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhelIleY 69
232 GGAGCCAAAGGGCCGCTGCACACCCATCATCTCTCGGAGGGGTACATCTC 281
69  sAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsnL 86
282 CAAGGACTACAAGGCAGAGCAACCTCACCACCTCCCGAGAGAGCGCA 331
86  euPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyr 102
332 CCTTCGTGATGGACATTGGCCACCTGACCCGCGGTGACTCTGGGCTTAC 381
103 AlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValTh 119
382 AAGTGTGGTCTGGGCATTAGCAC...CGAGGCCCTGCTTTTGTGATGTGAG 428
119 rLeuAsnVal 122
429 CCTGGAGGTC 438

```

CDS

LOCUS S43435S03 345 bp DNA PRI 08-MAY-1993
 DEFINITION transmembrane secretory component [human, leukocytes, Genomic, 345
 nt, segment 3 of 11].
 ACCESSION S43441
 VERSION S43441.1 GI:255089
 KEYWORDS 3 of 11
 SEGMENT human leukocytes.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 345)
 AUTHORS Krajci,P., Kvale,D., Tasken,K. and Brandtzaeg,P.
 TITLE Molecular cloning and exon-intron mapping of the gene encoding
 human transmembrane secretory component (the poly-Ig receptor)
 JOURNAL Eur J Immunol. 22 (9), 2309-2315 (1992)
 MEDLINE 92387236
 REMARK GenBank staff at the National Library of Medicine created this
 entry [NCBI gibbsq 113232] from the original journal article.
 This sequence comes from Fig. 2.

FEATURES
 source Location/Qualifiers
 1..345
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 78 a 102 c 97 g 68 t
 ORIGIN

alignment_scores:
 Quality: 182.00 Length: 115
 Ratio: 2.247 Gaps: 5
 Percent Similarity: 70.435 Percent Identity: 37.391

alignment_block:

US-09-135-238b-2 x S43435S03 ..

Align seg 1/1 to: S43435S03 from: 1 to: 345

12 ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGlycl 28
 |||:::|||||
 18 CCATATTGGT.....CCCGAG...GAGGTGAATAGTGT 49
 28 uLeuGlySerValThrIleLysCysProLeuPro.....G 41
 |||:::|||||
 50 GGAAGGTAACCTCAGTGTCCATCAGCTGCTACTACCCACCCACCTCTGTCA 99
 :: |||:::|||||
 41 luMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThr 57
 :: |||:::|||||
 100 ACCGCGCACCCCGGAAGTACTGGTCCGCGCAGGAGCTAGAGGTGGC... 146
 58 CysGlyThrValValSerThrThrAsnPhelLysAlaGluTyrLyscl 74
 ||| |||:::|||||
 147 TGCATAACCTCATCTCTCGGAGGCTACGTCCTCAGCAATATGCGAGG 196
 74 YargValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluV 91
 ||| |||:::|||||
 197 CAGGCTAACCTCACCAACTTCCCGAGAGCGCACATTGTGGTGAACA 246
 91 alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107
 :: |||:::|||||
 247 TTGCCAGCTAGCCAGGATGACTCCGGCGGCTACAAGTGTGGCTGGGC 296
 108 MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
 :: |||:::|||||
 297 ATCAATAGC...CGAGGCTGTCTTTGATGTACGCTGGAGGTC 338

seq_name: gb_pat:A52094

seq_documentation_block:

LOCUS A52094 1936 bp DNA PAT 11-MAR-1997
 DEFINITION Sequence 4 from Patent WO9618734.
 ACCESSION A52094
 VERSION A52094.1 GI:2304706

KEYWORDS .
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 1936)
 AUTHORS Van,O.J., Asselbergs,F.A., Leist,C. and Gruetter,M.
 TITLE PRODUCTION OF RECOMBINANT SECRETORY COMPONENT
 JOURNAL Patent: WO 9618734-A 4 20-JUN-1996;
 CIBA GEIGY AG (CH)
 COMMENT Other publication AU 4302796 960703.
 FEATURES Location/Qualifiers
 source 1..1936
 /organism="unidentified"
 /db_xref="taxon:32644"
 mat_peptide 16..1920
 BASE COUNT 451 a 521 c 583 g 381 t
 ORIGIN

alignment_scores:
 Quality: 182.00 Length: 115
 Ratio: 2.247 Gaps: 5
 Percent Similarity: 70.435 Percent Identity: 37.391

alignment_block:

US-09-135-238b-2 x A52094 ..

Align seg 1/1 to: A52094 from: 1 to: 1936

12 ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGlycl 28
 |||:::|||||
 76 CCATATTGGT.....CCCGAG...GAGGTGAATAGTGT 107
 28 uLeuGlySerValThrIleLysCysProLeuPro.....G 41
 |||:::|||||
 108 GGAAGGTAACCTCAGTGTCCATCAGCTGCTACTACCCACCCACCTCTGTCA 157
 41 luMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThr 57
 :: |||:::|||||
 158 ACCGCGCACCCCGGAAGTACTGGTCCGCGCAGGAGCTAGAGGTGGC... 204
 58 CysGlyThrValValSerThrThrAsnPhelLysAlaGluTyrLyscl 74
 ||| |||:::|||||
 205 TGCATAACCTCATCTCTCGGAGGCTACGTCCTCAGCAATATGCGAGG 254
 74 YargValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluV 91
 ||| |||:::|||||
 255 CAGGCTAACCTCACCAACTTCCCGAGAGCGCACATTGTGGTGAACA 304
 91 alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107
 :: |||:::|||||
 305 TTGCCAGCTAGCCAGGATGACTCCGGCGGCTACAAGTGTGGCTGGGC 354
 108 MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
 :: |||:::|||||
 355 ATCAATAGC...CGAGGCTGTCTTTGATGTACGCTGGAGGTC 396

.....

OM of: US-09-135-238B-2_COPY_18_253 to: GenEmbl.* out_format : pfs
Date: Sep 12, 2000 6:53 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=Cgnt_1/USPTO.spool/US09135238/runat_24082000_091323_17880/app_query.fasta_1.1282
-DB=GenEmbl -OFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.000 -OGAPEXT=0.000 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40 cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=1000000
-USER=US09135238_CGNT_1_5120 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-135-238B-2_COPY_18_253
Query length: 236
Database: GenEmbl.*
Database sequences: 972840
Database length: 892348106
Search time (sec): 2784.300000

score_list:	Strd Orig	ZScore	Escore	Len	Documentation
gb_pr3:AF057557	+ 1254.00	1700.48	1.3e-86	1339	! AF057557 Homo sapiens anti-Fas
gb_hg911:AC023534	+ 629.50	806.66	7.8e-37	221365	! AC023534 Homo sapiens clone
gb_hg922:AC027719	- 252.00	304.59	7.2e-09	58628	! AC027719 Homo sapiens chromos
gb_ro:MM06431	- 197.00	255.23	4.0e-06	3095	! U06431 Mus musculus B6/CBAF1J
gb_on:AF091137	+ 189.00	245.06	1.5e-05	2848	! AF091137 Trichosurus vulpecula
gb_ro:MMPIGR03	+ 188.00	260.34	2.1e-06	417	! U83428 Mus musculus polymeric i
gb_ro:MMPIGR3	+ 188.00	258.68	2.6e-06	505	! Y16525 Mus musculus p1gR gene,
gb_pat:EI5253	+ 188.00	223.50	0.0002	29392	! E15253 Mouse pig receptor gen
gb_ro:AB001489	+ 188.00	232.50	0.0002	29392	! AB001489 Mus musculus DNA for
gb_ro:RNIGR	+ 185.50	239.11	3.2e-05	3269	! X15741 Rat mRNA for polymeric
gb_on:AC032195	+ 181.00	235.44	5.1e-05	2461	! AB032195 Sus scrofa mRNA for p
gb_pr2:S43435S03	+ 180.50	251.77	6.3e-06	345	! S43441 transmembrane secretory
gb_pat:A52094	+ 180.50	236.84	4.3e-05	1936	! A52094 Sequence 4 from Patent
gb_pat:A52093	+ 180.50	236.43	4.5e-05	2031	! A52093 Sequence 3 from Patent
gb_pat:S62403	+ 180.50	234.96	5.4e-05	2405	! A52091 Sequence 1 from Patent
gb_pr2:S62403	+ 180.50	233.35	6.7e-05	2898	! S62403 transmembrane secretory
gb_pr1:HSPIR	+ 180.50	233.29	6.7e-05	2919	! X73079 Homo sapiens encoding p
gb_on:BTPIGR1	+ 179.50	230.04	0.0001	3630	! X81371 B.taurus mRNA for bovin
gb_on:BOVPIGR	+ 179.00	229.61	0.0001	3527	! L04797 Bos taurus polymeric in
gb_pat:EI5470	+ 175.50	230.15	0.0001	1911	! E15470 cDNA encoding immunity
gb_on:OCIGRECR	+ 166.50	212.63	0.0010	3517	! X00412 Rabbit mRNA for poly-im
gb_hg911:AC023534	- 166.00	176.09	0.1035	221365	! AC023534 Homo sapiens clone
gb_ro:MMPIGR5	+ 146.50	203.10	0.0032	456	! Y16527 Mus musculus p1gR gene,
gb_pr1:HUMIGRPOLY	+ 143.00	183.52	0.0399	2526	! M24559 Human poly-Ig receptor
gb_ro:MMPIGR05	+ 137.00	189.72	0.0180	381	! U83430 Mus musculus polymeric i
gb_pr2:S43435S05	+ 132.50	186.78	0.0263	333	! S43443 transmembrane secretory
gb_pr2:S43435S04	+ 130.00	177.49	0.0864	657	! S43442 transmembrane secretory
gb_pr2:HS2428323	+ 124.00	165.97	0.3791	969	! AJ238323 Homo sapiens mRNA for
gb_pr2:HS2428323	+ 124.00	165.96	0.3795	970	! AJ010101 Homo sapiens mRNA for
gb_pr2:HS2428323	+ 124.00	160.34	0.7805	1857	! AJ224864 Homo sapiens mRNA for
gb_hg929:AC060797	- 121.00	126.24	61.86	59506	! AC060797 Homo sapiens chromos
gb_hg915:AC016888	- 121.00	116.89	205.24	175260	! AC016888 Homo sapiens chromid
gb_pr3:AF043724	+ 117.50	153.70	1.83	1440	! AF043724 Homo sapiens hepatic
gb_on:RABTCBXB	+ 114.50	150.17	2.88	1351	! M13895 Rabbit T-cell receptor
gb_ro:AF035963	+ 111.50	140.62	9.78	2539	! AF035963 Rattus norvegicus kid
gb_pr1:HSNCAMX1	+ 111.00	123.85	84.04	16288	! Z29373 H.sapiens gene for neu
gb_pat:AF000062	+ 110.50	99.47	1.9e+03	251700	! AP000062 Aeropyrum pernix ge
gb_hg922:AC027719	+ 109.50	110.73	452.59	58628	! AC027719 Homo sapiens chromos
gb_hg915:AC013305	- 109.50	101.22	1.5e+03	175799	! AC013305 Homo sapiens clone
gb_pr2:HUMTCBZY	+ 109.00	144.67	5.82	1074	! M15564 Human T-cell receptor r

gb_pr2:HUMTCBZAL + 108.50 145.18 5.45 936 ! L34734 Homo sapiens T-cell r
gb_hg928:AC021643 - 108.00 96.02 3.0e+03 253327 ! AC021643 Mus musculus chr
gb_ro:MMPIGR04 + 107.50 145.84 5.01 741 ! U83429 Mus musculus polymeri
gb_ro:MMPIGR4 + 107.50 145.22 5.42 796 ! Y16526 Mus musculus p1gR gen
gb_ov:D85084 + 107.50 130.43 36.17 4398 ! D85084 Cynops pyrrhogaster

seq_name: gb_pr3:AF057557

seq documentation block:
LOCUS AF057557 1339 bp mRNA PRI 04-JUN-1998
DEFINITION Homo sapiens anti-Fas-induced apoptosis (TOSO) mRNA, complete cds.
ACCESSION AF057557
VERSION AF057557.1 GI:3169292
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1339)
AUTHORS Hitoshi,Y., Lorens,J., Kitada,S.I., Fisher,J., LaBarge,M.,
Ring,H.Z., Francke,U., Reed,J.C., Kinoshita,S. and Nolan,G.P.
TITLE Toso, a cell surface, specific regulator of Fas-induced apoptosis
in T cells
JOURNAL Immunity 8 (4), 461-471 (1998)
MEDLINE 98246048
REFERENCE 2 (bases 1 to 1339)
AUTHORS LaBarge,M. and Hitoshi,Y.
TITLE Direct Submission
JOURNAL Submitted (03-APR-1998) Molecular Pharmacology, Stanford
University, 300 Pasteur Dr., Palo Alto, CA 94304, USA

FEATURES
Location/Qualifiers
1..1339
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q31-q32"
/cell_line="Jurkat"
/cell_type="T cell"
/note="TOSO.2 cDNA; cloned by PCR using primers designed
from the original TOSO.1 cDNA sequence"
1..1339
/gene="TOSO"
20..1192
/gene="TOSO"
/note="transmembrane protein; Ig superfamily member"
/codon_start=1
/product="anti-Fas-induced apoptosis"
/protein_id="AAC18830.1"
/db_xref="GI:3169293"
/translation="MDFWMLPFLPVSGALRILPEVKVEGELGGSVTIKCPLPEMHV
RIYLCKEMAGSGCTGVVSTNFIKAEYKGRVTLKQYPRKNLFLVEVTLQTESDGVY
ACGAGMNDRGKTKVTLNVHSEYEPWEQPMPTPKWHLPLYLFQMPAYASSKFEV
TRVTPPAQRGKVPVYHSSPTQITHRPVRASSVAGDKPRTFLPFTASKISALEG
LKPTPTPSYNNHTRHLRQALDYSQSGREGQGHILPTILGLFLLALLGLVWKRVA
ERRKALSRARLEAYRMRALDESQRPSRPNISYACPRRACDAAGTCEAP
VPCGAPLEPPAPLQVSESPWLEAPSLKTSCEYSLYHQHPAMMEDSDSDDDYINWPA"
127
variation
/gene="TOSO"
/note="occurs in TOSO.1 cDNA"
/replace="a"
1015
/gene="TOSO"
/note="occurs in TOSO.1 cDNA"
/replace="g"
BASE COUNT 309 a 432 c 340 g 258 t
ORIGIN

alignment_scores:
Quality: 1254.00 Length: 236
Ratio: 5.314 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-135-238B-2_Copy_18_253 x AF057557

Align seg 1/1 to: AF057557 from: 1 to: 1339

```
1  ArgIleLeuProGluValLysValGluGlyLeuGlySerValTh 17
|||||
71  AGGATCTCCAGAGTAAGGTAGAGGGGAGCTGGCGGATCGATTAC 120
|||||
17  rIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 34
|||||
121  CATCAGTGCCACCTCTCGAATGCATGTGAGGATATATCTGTCCGGG 170
|||||
34  luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
|||||
171  AGATGGCTGGATCTGGAACATGTGTACCGTGGTATCCACCACTTC 220
|||||
51  ileLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67
|||||
221  ATCAGGCAGATACAGGGCGGAGTACTCTGAAGCAATACCCAGCAA 270
|||||
67  sasnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84
|||||
271  GAATCTGTCTCTAGTGGAGGTACACAGCTGACAGAAAGTGACACGGAG 320
|||||
84  alTyrAlaCysGlyValAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
|||||
321  TCTATGCTGGCGGAGCGGATGACACAGACCGGGGAGAGACCCAGAAA 370
|||||
101  ValThrLeuAsnValHisSerGluTyrGluProSerTrpGluGluInPr 117
|||||
371  GTCCACCTGAATGTCCACAGTGAATACAGGCCATCATGGGAGAGACGCC 420
|||||
117  oMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMetP 134
|||||
421  AATGCCCTGAGACTCCAAATGGTTTCATCTGCCCTATTGTTCCAGATGC 470
|||||
134  roAlaTyrAlaSerSerLysPheValThrArgValThrThrProAla 150
|||||
471  CTCATATGCCAGTCTCTCCAAATTCGTAACCCAGAGTTACACACAGCT 520
|||||
151  GlnArgGlyLysValProProValHisHisSerSerProThrThrGlnIl 167
|||||
521  CAAAGGGGCAAGGTCCTCCAGTTCACCACTCTCTCCACCAACCAAT 570
|||||
167  eThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLysP 184
|||||
571  CACCCACCGCCCTCGAGTGTCCAGAGCATCTTCAGTAGCAGGTGACAAAG 620
|||||
184  roArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuGlu 200
|||||
621  CCGGAACCTCTCTGCCATCCACTACAGCTCAAAAATCTCAGCTCTGGAG 670
|||||
201  GlyLeuLeuLysProGlnThrProSerTyrAsnHisHisThrArgLeuHi 217
|||||
671  GGCTCTCTCAAGCCCGCAGAGCCCGAGCTACAAACCCACACACAGGCTGCA 720
|||||
217  sArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGlyGlnG 234
|||||
721  CAGGACAGAGCACTGGCACTATGGCTACAGTCTGGGAGGAGGCCAAG 770
|||||
234  lyPheHis 236
|||||
771  GATTTTCAAC 778
```

seq_name: gb_htg11:AC023534

seq_documentation_block:

LOCUS AC023534 221365 bp DNA HTG 02-MAR-2000
DEFINITION Homo sapiens clone RP11-462N18, *** SEQUENCING IN PROGRESS ***, 61
unorderd pieces.
ACCESSION AC023534
VERSION AC023534.2 GI:7143452

KEYWORDS
SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

HTG; HTGS_PHASE1.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 221365)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome, clone RP11-462N18

Unpublished

2 (bases 1 to 221365)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavskiy,L.,

Bouckgaert,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,

Choepeil,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,

DeArelano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,

Fenestor,J., Ferrel,P., FitzHugh,W., Forrest,C., Gage,D.,

Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,

Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,

Mcwan,P., McGurk,A., McKernan,K., McPheeters,K., Meldrim,J.,

Mense,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J.,

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivari,T.M.,

Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,

Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,

Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,

Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,

Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and

Zody,M.

Direct Submission

Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 2, 2000 this sequence version replaced gi:6978210.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5244

Center clone name: 462_N_18

* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1108: contig of 1108 bp in length
* 1109 1208: gap of 100 bp
* 1209 2292: contig of 1084 bp in length
* 2293 2392: gap of 100 bp
* 2393 3673: contig of 1281 bp in length
* 3674 3773: gap of 100 bp
* 3774 4947: contig of 1174 bp in length
* 4948 5047: gap of 100 bp
* 5048 6104: contig of 1057 bp in length
* 6105 6204: gap of 100 bp
* 6205 7259: contig of 1055 bp in length
* 7260 7359: gap of 100 bp
* 7360 8667: contig of 1308 bp in length
* 8668 8767: gap of 100 bp
* 8768 10586: contig of 1799 bp in length
* 10567 10666: gap of 100 bp
* 10667 12173: contig of 1507 bp in length
* 12174 12273: gap of 100 bp

```
* 12274 13643: contig of 1370 bp in length
* 13644 13743: gap of 100 bp
* 13744 15215: contig of 1472 bp in length
* 15216 15315: gap of 100 bp
* 15316 16609: contig of 1294 bp in length
* 16610 16709: gap of 100 bp
* 16710 18085: contig of 1376 bp in length
* 18086 18185: gap of 100 bp
* 18186 19668: contig of 1483 bp in length
* 19669 19768: gap of 100 bp
* 19769 22083: contig of 2315 bp in length
* 22084 22183: gap of 100 bp
* 22184 24052: contig of 1869 bp in length
* 24053 24152: gap of 100 bp
* 24153 25902: contig of 1750 bp in length
* 25903 26002: gap of 100 bp
* 26003 27639: contig of 1637 bp in length
* 27640 27739: gap of 100 bp
* 27740 29377: contig of 1638 bp in length
* 29378 29477: gap of 100 bp
* 29478 30910: contig of 1433 bp in length
* 30911 31010: gap of 100 bp
* 31011 31225: contig of 215 bp in length
* 31226 31325: gap of 100 bp
* 31326 32890: contig of 1565 bp in length
* 32891 32990: gap of 100 bp
* 32991 34812: contig of 1822 bp in length
* 34813 34912: gap of 100 bp
* 34913 36817: contig of 1905 bp in length
* 36818 36917: gap of 100 bp
* 36918 38575: contig of 1658 bp in length
* 38576 38675: gap of 100 bp
* 38676 41309: contig of 2634 bp in length
* 41310 41409: gap of 100 bp
* 41410 43344: contig of 1935 bp in length
* 43345 43444: gap of 100 bp
* 43445 45853: contig of 2409 bp in length
* 45854 45953: gap of 100 bp
* 45954 48434: contig of 2481 bp in length
* 48435 48534: gap of 100 bp
* 48535 51026: contig of 2492 bp in length
* 51027 51126: gap of 100 bp
* 51127 54369: contig of 3243 bp in length
* 54370 54469: gap of 100 bp
* 54470 56681: contig of 2192 bp in length
* 56682 56781: gap of 100 bp
* 56782 59097: contig of 2336 bp in length
* 59098 59197: gap of 100 bp
* 59198 61551: contig of 2354 bp in length
* 61552 61651: gap of 100 bp
* 61652 64257: contig of 2606 bp in length
* 64258 64357: gap of 100 bp
* 64358 67511: contig of 3154 bp in length
* 67512 67611: gap of 100 bp
* 67612 70329: contig of 2718 bp in length
* 70330 70429: gap of 100 bp
* 70430 73572: contig of 3143 bp in length
* 73573 73672: gap of 100 bp
* 73673 77203: contig of 3531 bp in length
* 77204 77303: gap of 100 bp
* 77304 80599: contig of 3296 bp in length
* 80600 80699: gap of 100 bp
* 80700 84046: contig of 3347 bp in length
* 84047 84146: gap of 100 bp
* 84147 88357: contig of 4211 bp in length
* 88358 88457: gap of 100 bp
* 88458 92454: contig of 3997 bp in length
* 92455 92554: gap of 100 bp
* 92555 96893: contig of 4338 bp in length
* 96893 96992: gap of 100 bp
* 96993 101334: contig of 4342 bp in length
* 101335 101434: gap of 100 bp
* 101435 105438: contig of 4004 bp in length
```

```
* 105439 105538: gap of 100 bp
* 105539 110413: contig of 4875 bp in length
* 110414 110513: gap of 100 bp
* 110514 115994: contig of 5481 bp in length
* 115995 116094: gap of 100 bp
* 116095 120447: contig of 4353 bp in length
* 120448 120547: gap of 100 bp
* 120548 124483: contig of 3936 bp in length
* 124484 124583: gap of 100 bp
* 124584 130249: contig of 5666 bp in length
* 130250 130349: gap of 100 bp
* 130350 135462: contig of 5113 bp in length
* 135463 135562: gap of 100 bp
* 135563 139847: contig of 4285 bp in length
* 139848 139947: gap of 100 bp
* 139948 146731: contig of 6784 bp in length
* 146732 146831: gap of 100 bp
* 146832 151469: contig of 4638 bp in length
* 151470 151569: gap of 100 bp
* 151570 159901: contig of 8332 bp in length
* 159902 160001: gap of 100 bp
* 160002 168308: contig of 8307 bp in length
* 168309 168408: gap of 100 bp
* 168409 176195: contig of 7787 bp in length
* 176196 176295: gap of 100 bp
* 176296 189303: contig of 13008 bp in length
* 189304 189403: gap of 100 bp
* 189404 202135: contig of 12732 bp in length
* 202136 202235: gap of 100 bp
* 202236 221365: contig of 19130 bp in length.

FEATURES
      Location/Qualifiers
      source
        1..221365
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="RP11-462N18"
        /clone_lib="RPCI-11 Human Male BAC"
      misc_feature
        1..1108
        /note="assembly_fragment"
      misc_feature
        1209..2292
        /note="assembly_fragment"
      misc_feature
        2393..3673
        /note="assembly_fragment"
      misc_feature
        3774..4947
        /note="assembly_fragment"
      misc_feature
        5048..6104
        /note="assembly_fragment"
      misc_feature
        6205..7259
        /note="assembly_fragment"
      misc_feature
        7360..8667
        /note="assembly_fragment"

alignment_scores:
      Quality: 629.50      Length: 476
      Ratio: 3.459      Gaps: 6
      Percent Similarity: 38.235      Percent Identity: 34.244

alignment_block:
US-09-135-238B-2_COPY_18_253 x AC023534 ..
Align seg 1/1 to: AC023534 from: 1 to: 221365
1 Argf1eLeuProGluValLysValGluGlyGluLeuGlyGlySerValth 17
|||||
33190 AGATCTCCCAAGTAAGGTAGAGGGGAGCTGGCGGATCATCTTAC 33239
|||||
17 r1eLysCysProLeuProGluMethisValArgilletyrLeuCysArg 34
|||||
33240 CATCAAGTCCCACTTCCCTGAAATGCATGTGAGGATATATCTGTGCGG 33289
|||||
34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
|||||
33290 AGATGGCTGGATCTGGAACATGTGTACGTTGATCCACCAACTTC 33339
|||||
51 l1eLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgly 67
|||||
```


Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7449
Center clone name: 258_P_9

* NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 671: contig of 671 bp in length
* 672 771: gap of 100 bp
* 772 1469: contig of 698 bp in length
* 1470 1569: gap of 100 bp
* 1570 2261: contig of 692 bp in length
* 2262 2361: gap of 100 bp
* 2362 3051: contig of 690 bp in length
* 3052 3151: gap of 100 bp
* 3152 3816: contig of 665 bp in length
* 3817 3916: gap of 100 bp
* 3917 4597: contig of 681 bp in length
* 4598 4697: gap of 100 bp
* 4698 5395: contig of 698 bp in length
* 5396 5495: gap of 100 bp
* 5496 6185: contig of 690 bp in length
* 6186 6285: gap of 100 bp
* 6286 6969: contig of 684 bp in length
* 6970 7069: gap of 100 bp
* 7070 7780: contig of 711 bp in length
* 7781 7880: gap of 100 bp
* 7881 8590: contig of 710 bp in length
* 8591 8690: gap of 100 bp
* 8691 9372: contig of 682 bp in length
* 9373 9472: gap of 100 bp
* 9473 10162: contig of 690 bp in length
* 10163 10262: gap of 100 bp
* 10263 10944: contig of 682 bp in length
* 10945 11044: gap of 100 bp
* 11045 11747: contig of 703 bp in length
* 11748 11847: gap of 100 bp
* 11848 12547: contig of 700 bp in length
* 12548 12647: gap of 100 bp
* 12648 13336: contig of 689 bp in length
* 13337 13436: gap of 100 bp
* 13437 14103: contig of 667 bp in length
* 14104 14203: gap of 100 bp
* 14204 14872: contig of 669 bp in length
* 14873 14972: gap of 100 bp
* 14973 15682: contig of 710 bp in length
* 15683 15782: gap of 100 bp
* 15783 16491: contig of 709 bp in length
* 16492 16591: gap of 100 bp
* 16592 17290: contig of 699 bp in length
* 17291 17390: gap of 100 bp
* 17391 18088: contig of 698 bp in length
* 18089 18188: gap of 100 bp
* 18189 18883: contig of 695 bp in length
* 18884 18983: gap of 100 bp
* 18984 19671: contig of 688 bp in length
* 19672 19771: gap of 100 bp
* 19772 20413: contig of 642 bp in length

* 20414 20513: gap of 100 bp
* 20514 21196: contig of 683 bp in length
* 21197 21296: gap of 100 bp
* 21297 21992: contig of 696 bp in length
* 21993 22992: gap of 100 bp
* 22093 22804: contig of 712 bp in length
* 22805 22904: gap of 100 bp
* 22905 23604: contig of 700 bp in length
* 23605 23704: gap of 100 bp
* 23705 24402: contig of 698 bp in length
* 24403 24502: gap of 100 bp
* 24503 25191: contig of 689 bp in length
* 25192 25291: gap of 100 bp
* 25292 25974: contig of 683 bp in length
* 25975 26074: gap of 100 bp
* 26075 26769: contig of 695 bp in length
* 26770 26869: gap of 100 bp
* 26870 27572: contig of 703 bp in length
* 27573 27672: gap of 100 bp
* 27673 28373: contig of 701 bp in length
* 28374 28473: gap of 100 bp
* 28474 29165: contig of 692 bp in length
* 29166 29265: gap of 100 bp
* 29266 29956: contig of 691 bp in length
* 29957 30056: gap of 100 bp
* 30057 30741: contig of 685 bp in length
* 30742 30841: gap of 100 bp
* 30842 31519: contig of 678 bp in length
* 31520 31619: gap of 100 bp
* 31620 32323: contig of 704 bp in length
* 32324 32423: gap of 100 bp
* 32424 33110: contig of 687 bp in length
* 33111 33210: gap of 100 bp
* 33211 33902: contig of 692 bp in length
* 33903 34002: gap of 100 bp
* 34003 34708: contig of 706 bp in length
* 34709 34808: gap of 100 bp
* 34809 35500: contig of 692 bp in length
* 35501 35600: gap of 100 bp
* 35601 36300: contig of 700 bp in length
* 36301 36400: gap of 100 bp
* 36401 37097: contig of 697 bp in length
* 37098 37197: gap of 100 bp
* 37198 37900: contig of 703 bp in length
* 37901 38000: gap of 100 bp
* 38001 38695: contig of 695 bp in length
* 38696 38795: gap of 100 bp
* 38796 39487: contig of 692 bp in length
* 39488 39587: gap of 100 bp
* 39588 40291: contig of 704 bp in length
* 40292 40391: gap of 100 bp
* 40392 41101: contig of 710 bp in length
* 41102 41201: gap of 100 bp
* 41202 41886: contig of 685 bp in length
* 41887 41986: gap of 100 bp
* 41987 42686: contig of 700 bp in length
* 42687 42786: gap of 100 bp
* 42787 43484: contig of 698 bp in length
* 43485 43584: gap of 100 bp
* 43585 44276: contig of 692 bp in length
* 44277 44376: gap of 100 bp
* 44377 45078: contig of 702 bp in length
* 45079 45178: gap of 100 bp
* 45179 45869: contig of 691 bp in length
* 45870 45969: gap of 100 bp
* 45970 46664: contig of 695 bp in length
* 46665 46764: gap of 100 bp
* 46765 47456: contig of 692 bp in length
* 47457 47556: gap of 100 bp
* 47557 48248: contig of 693 bp in length
* 48250 48349: gap of 100 bp
* 48350 49042: contig of 693 bp in length
* 49043 49142: gap of 100 bp

```

* 49143 49843: contig of 701 bp in length
* 49844 49943: gap of 100 bp
* 49944 50640: contig of 697 bp in length
* 50641 50740: gap of 100 bp
* 50741 51438: contig of 698 bp in length
* 51439 51538: gap of 100 bp
* 51539 52331: contig of 693 bp in length
* 52332 52331: gap of 100 bp
* 52332 53020: contig of 689 bp in length
* 53021 53120: gap of 100 bp
* 53121 53857: contig of 737 bp in length
* 53858 53957: gap of 100 bp
* 53958 54652: contig of 695 bp in length
* 54653 54752: gap of 100 bp

```

alignment_scores:

```

Quality: 252.00 Length: 48
Ratio: 5.250 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

```
US-09-135-238B-2_COPY_18_253 x AC027719/rev ..
```

```
Align seg 1/1 to reverse of: AC027719 from: 1 to: 58628
```

```

1 ArgileLeuProGluValLysValGluGlySerValTh 17
|||||
40580 AGGATCTCCAGAACTAAAGTAGAGGGGAGCTGGGGGATCAGTTAC 40531
|||||

17 rtleYcysProLeuProGluMetHisValArgIleTyrLeuCysArg 34
|||||
40530 CATCAATGCCCACTCTCTGAAATGCATGTGAGGATATATCTGTGCGGG 40481
|||||

34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThr 48
|||||
40480 AGATGGCTGGATCTGGACATGTGGTACCGTGTATCCACCACC 40437
|||||

```

```
seq_name: gb_ro:MMU06431
```

```

seq_documentation_block:
LOCUS MMU06431 3095 bp mRNA ROD 26-MAY-1995
DEFINITION Mus musculus B6/CBAF1J polymeric immunoglobulin receptor mRNA,
complete cds.

```

```

ACCESSION U06431
VERSION U06431.1 GI:458421
KEYWORDS
SOURCE mouse.

```

ORGANISM

```

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE 1 (bases 1 to 3095)
AUTHORS Piskurich,J.F., Blanchard,M.H., Youngman,K.R., France,J.A. and Kaetzel,C.S.

```

```

TITLE Molecular cloning of the mouse polymeric Ig receptor. Functional
regions of the molecule are conserved among five mammalian species
J. Immunol. 154 (4), 1735-1747 (1995)

```

JOURNAL

MEDLINE

```
2 (sites)
Piskurich,J.F., Hsieh,M., Youngman,K.R. and Kaetzel,C.S.

```

REFERENCE

```

AUTHORS Piskurich,J.F.
TITLE Molecular cloning of mouse polymeric Ig receptor
J. Immunol. 150, 38A-38A (1993)

```

JOURNAL

REFERENCE

```

3 (bases 1 to 3095)
Piskurich,J.F.

```

AUTHORS

```

TITLE Molecular Cloning and Regulation of the Polymeric Immunoglobulin
Receptor
Thesis (1994) Pathology, Case Western Reserve University

```

JOURNAL

REFERENCE

```

4 (bases 1 to 3095)
Piskurich,J.F.

```

AUTHORS

```

TITLE Direct Submission
Submitted (08-FEB-1994) Janet F. Piskurich, Case Western Reserve
University, Pathology, Biomedical Research Building, Cleveland, OH
44060, USA

```

JOURNAL

FEATURES

```

Location/Qualifiers

```

```

source
1. .3095
/organism="Mus musculus"
/strain="B6/CBAF1J (C57 Black 6 x CBA)"
/db_xref="taxon:10090"
/clone="G20"
/clone_lib="catalog #935302, Stratagene Cloning Systems,
La Jolla, CA"
/sex="female"
/tissue_type="liver"
<1. .84
85. .138
85. .2400
/codon_start=1
/function="binds and transports polymeric immunoglobulin"
/product="polymeric immunoglobulin receptor"
/protein_id="AA67440.1"
/db_xref="GI:458422"
/translation="MRLYLFTLLTVFSGVSTKSPIFGPQEVSSIEGDSVSICTYYPD
TSVNRHTRKYWCQAGSMCTTLLISSGYSKEYSGRANLINFENNTFVINTQLTQ
DDTGYKCGCLGTSNRGLSDVSEVSVQPELPSDTHVTKDGRNVTIECPFKRENVP
SKKSLCKTNQSCSLVIDSTKVNPSYIGRAKLFMKGTDLTFVYVNIHSLTHNDAGLY
ICQAGEPSADKNVDLOVLAPEPELLYKDLRSVTLFECDLGREVANEAKVLCRMNKE
TCDVIINTLGRKDPDPFGRILITPKDDNGRFSVLITGLRKEDAGHVQCGASHSLPQE
GWPLOTWOLFVNEESTIPNRSVGVGTGGSVALACPNPKESSLSIAKYNCRWBGDNG
HCPALVGTQAQVQEEYEGRLALFDOPNGHYTVLLAQLTDEAGFYWCLNGSDGRHIT
TIELQVAETREPENLVEITPONATAVLGTEFTVVSCHTFCPAFTSQEKYCKWNSKNGCHIL
PSHDEGARQSVSDCDSSQLVSMTLNPVSRKDEGWYCGVKGQGTGYGETTAYIAVEE
ENRAGSDAGSADGQSRSSSKVLFSTLPLGLVLAVGAIAVWVAVRHRKKNVDRMSIS
SYRTDISMADPKNSRDIGGNDMGASPDQTQTVIEGKDEIVTTTCTAEPEESKAKR
SSKEADMAVSFAFLLOSSTIAAQVHDGPOEA"
139. .2397
mat_peptide
139. .2019
misc_feature
139. .2019
misc_feature
2020. .2088
/note="ectoplasmic region, binds polymeric immunoglobulin"
/note="transmembrane segment of the receptor"
2089. .2397
misc_feature
2401. .3095
/note="cytoplasmic tail of the receptor"
3' UTR
BASE COUNT 861 a 796 c 784 g 654 t
ORIGIN
alignment_scores:
Quality: 197.00 Length: 148
Ratio: 2.165 Gaps: 5
Percent Similarity: 61.486 Percent Identity: 32.432
alignment_block:
US-09-135-238B-2_COPY_18_253 x MMU06431 ..
Align seg 1/1 to: MMU06431 from: 1 to: 3095

```

```

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
||| |||||:||||| ||| |||||
181 GCGCAGCTCTGTTCCATCAGCTGCTACTACCAGACACCTCTGTCAACCG 230
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||||| ||| |||||: ||| ||||| |||
231 GCACACCGGAAATACTGTGCGCAGCAGGAGCC...AGCGGCATGTCCA 277
42 lyThrValValSerThrThrAsnPhelIleLysAlaGluTyrLysGlyArg 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||
278 CAACGCTCATCTTCAATATGCTACCTCTCCAAGGAGTATTTCAGGCAGA 327
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
:|||||: |||||: |||||: |||||: |||||: |||||: |||||
328 GCCAACCTCATCAACTTCCCGAGAGAACACACATTGTGATTAAACATTGA 377
75 rGlnLeuThrGluSerAspSerClyValTyrAlaCysGlyAlaGlyMeta 92
|||||:|||||:|||||:|||||: ||||| ||||| |||||
378 GCAGCTCACCCAGGACGACACTGGGAGCTACAGTGTGCGCTGGGT...A 424

```


REFERENCE 2 (bases 1 to 417)
 AUTHORS Martin,M.G., Gutierrez,E.M., Lam,J.T., Li,T.W.H. and Wang,J.
 TITLE Direct Submission
 JOURNAL Submitted (31-DEC-1996) Pediatrics, UCLA School of Medicine, 10833 Le Conte Ave, Los Angeles, CA 90095, USA

FEATURES
 source Location/Qualifiers
 1. .417
 /organism="Mus musculus"
 /strain="129SVJ"
 /db_xref="taxon:10090"
 exon 43. .387
 /gene="pIgr"
 /number=3
 BASE COUNT 105 a 120 c 102 g 90 t
 ORIGIN

alignment_scores:
 Quality: 188.00 Length: 97
 Ratio: 2.648 Gaps: 3
 Percent Similarity: 73.196 Percent Identity: 40.206

alignment_block:
 US-09-135-238b-2_copy_18_253 x MMPIGR03 ..

Align seg 1/1 to: MMPIGR03 from: 1 to: 417

```

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
||| |||||:||||| ||| |||||:
96 GGGGACTCTGTTCCATCAGTGCTACTACCCAGACACCTCTGTCAACCG 145
25 tHtSValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||| |||||:||||| ||| |||||: |||
146 GCACACCCGGAAATCTGGTGGCCGACAAAGAGCC...AGCGCATGTGCA 192
42 lyThrValValSerThrThrAsnPhelLysAlaGluTyrLysGlyArg 58
||| |||||:||||| ||| |||||: |||
193 CAACGCTCATCTTCAATGGTCTACTCCAAGGAGTATTTCAGGCAGA 242
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||||:|||||:|||||:|||||:|||||:|||||:
243 GCCACCTCATCACTTCCAGAGAACACACATTTGTGATTACATTGA 292
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
|||||:|||||:|||||:||||| ||| |||||: |||
293 GCAGCTCACCCAGACACACTGGGAGCTACAAGTGTGGCTGGGT...A 339
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 105
|||||:|||||: ||| |||||:|||||
340 CCAGTAACCGAGGCTCTCTTCGATGTCAGCCTGGAGGTC 380

```

seq_name: gb_ro:MMPIGR3

seq_documentation_block:
 LOCUS MMPIGR3 505 bp DNA ROD 09-APR-1998
 DEFINITION Mus musculus pIgr gene, exon 3.
 ACCESSION Y16525
 VERSION Y16525.1 GI:3046750
 KEYWORDS pIgr gene; polymeric immunoglobulin receptor.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 505)
 AUTHORS de Groot,N., Vollebregt,E., Lee,S.H., Verbeet,M.P. and de Boer,H.A.
 TITLE Molecular cloning and exon-intron organization of the gene encoding the murine polymeric immunoglobulin receptor
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 505)
 AUTHORS de Groot,N.
 TITLE Direct Submission
 JOURNAL Submitted (11-FEB-1998) N. De Groot, Medical Biotechnology Dept., Leiden University, P.O. box 9502, 2300 RA Leiden, NETHERLANDS

FEATURES
 source Location/Qualifiers
 1. .505
 /organism="Mus musculus"
 /strain="129"
 /db_xref="taxon:10090"
 /tissue_type="liver"
 77. .421
 /gene="pIgr"
 77. .421
 /gene="pIgr"
 /label-ex3
 /usedin=Y16524:pig_cds
 /usedin=Y16524:pig_mrna
 BASE COUNT 130 a 136 c 127 g 112 t
 ORIGIN

alignment_scores:
 Quality: 188.00 Length: 97
 Ratio: 2.648 Gaps: 3
 Percent Similarity: 73.196 Percent Identity: 40.206

alignment_block:
 US-09-135-238b-2_copy_18_253 x MMPIGR3 ..

Align seg 1/1 to: MMPIGR3 from: 1 to: 505

```

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
||| |||||:||||| ||| |||||:
130 GGGGACTCTGTTCCATCAGTGCTACTACCCAGACACCTCTGTCAACCG 179
25 tHtSValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
|||||:|||||:|||||:||||| ||| |||||: |||
180 GCACACCCGGAAATCTGGTGGCCGACAAAGAGCC...AGCGCATGTGCA 226
42 lyThrValValSerThrThrAsnPhelLysAlaGluTyrLysGlyArg 58
|||||:|||||:|||||:|||||:|||||:|||||:
227 CAACGCTCATCTTCAATGGTCTACTCCAAGGAGTATTTCAGGCAGA 276
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||||:|||||:|||||:|||||:|||||:|||||:
277 GCCACCTCATCACTTCCAGAGAACACACATTTGTGATTACATTGA 326
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
|||||:|||||:|||||:||||| ||| |||||: |||
327 GCAGCTCACCCAGACACACTGGGAGCTACAAGTGTGGCTGGGT...A 373
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 105
|||||:|||||: ||| |||||:|||||
374 CCAGTAACCGAGGCTCTCTTCGATGTCAGCCTGGAGGTC 414

```

seq_name: gb_pat:E15253

seq_documentation_block:
 LOCUS E15253 29392 bp DNA PAT 28-JUL-1999
 DEFINITION Mouse pIgr receptor gene.
 ACCESSION E15253
 VERSION E15253.1 GI:5709936
 KEYWORDS JP 1998057066-A/1.
 SOURCE Mus sp..
 ORGANISM Mus sp.

REFERENCE 1 (bases 1 to 29392)
 AUTHORS Sato,T. and Kushi,A.
 TITLE MURINE PIG RECEPTOR PROTEIN GENE
 JOURNAL Patent: JP 1998057066-A 03-MAR-1998;
 YAKULT HONSHA CO LTD
 COMMENT OS Mus sp. (mouse)
 PN JP 1998057066-A/1
 PD 03-MAR-1998
 PF 19-AUG-1996 JP 1996217154
 PI SATO TAKASHI, KUSHIRO AKIRA

```
PC C12N15/09,C07H21/04,C07K14/705;
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FH source 1..29392
FH /organism="Mus sp."
FH /clone_lib="genomic DNA from female 129SVJ FT
FH mouse liver in
FH lambda fix vector".
FT Location/Qualifiers
FT 1..29392
FT /organism="Mus sp."
FT /db_xref="taxon:10095"
BASE COUNT 8318 a 6747 c 6514 g 7813 t
ORIGIN

alignment_scores:
  Quality: 188.00      Length: 97
  Ratio: 2.648         Gaps: 3
Percent Similarity: 73.196      Percent Identity: 40.206

alignment_block:
US-09-135-238B-2_COPY_18_253 x E15253 ..
Align seg 1/1 to: E15253 from: 1 to: 29392

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
||| |||||:|||| ||| |||||
18245 GCGGACTCTGTTTCCATCAGTCTACTACCGACACCTCTGTCAACCG 18294

25 thisValArgIleTyLeuCysArgGluMetAlaGlySerGlyThrCysG 42
|||||:|||| ||| |||||:|||| ||| |||||:||||
18295 GCACACCCGGAATACTGTGCGCAGCAGGAGCC...ACGGCAGTGTGA 18341

42 lyThrValValSerThrThrAsnPhelLeLysAlaGluTyrlLysGlyArg 58
|||||:||||:||||:||||:||||: |||||:|||||
18342 CAACGCTCATCTCTTCAAATGCTACCTCTCCAAGGAGTATTTCAGGCAGA 18391

59 valThrLeuLysGlnTyrrProargLysAsnLeuPheValIgluValth 75
|||||:|||| :|||||:||||:|||| |||||:||||:
18392 GCAACCTCATCAACTCCAGACACACACATTGTGATTAAACATTGA 18441

75 rGlnLeuThrGluSerAspSerGlyValTyrrAlaCysGlyAlaGlyMeta 92
|||||:||||:||||:||||: ||| |||||:||||
18442 GCAGCTACCCAGGAGCAGCACGTGGGAGCTACAGTGTGGCCTGGGT...A 18488

92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 105
:||||:||||| :||| |||||:|||||
18489 CCAGTAACCGAGGCGTGTCTTCGATGTCAGCGCTGAGGTC 18529

seq_name: gb_ro:AB001489

seq_documentation_block:
LOCUS AB001489 29392 bp DNA ROD 14-APR-2000
DEFINITION Mus musculus DNA for polyimmunoglobulin receptor, complete cds.
ACCESSION AB001489
VERSION AB001489.1 GI:2804245
KEYWORDS polyimmunoglobulin receptor.
SOURCE Mus musculus (strain:129SVJ) female liver DNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
  1 (sites)
  AUTHORS kushiro,A. and Sato,T.
  TITLE PolymERIC immunoglobulin receptor gene of mouse: sequence,
    structure and chromosomal location
  JOURNAL Gene 204 (1-2), 277-282 (1997)
  MEDLINE 98094283
  REFERENCE 2 (bases 1 to 29392)
  AUTHORS Sato,T.
  TITLE Direct Submission
```

```
JOURNAL Submitted (26-FEB-1997) to the DDBJ/EMBL/GenBank databases. Takashi
Sato, Yakult Central Institute for Microbiological Research,
Culture Collection Center; 1796 Yaho, Kunitachi-shi, Tokyo 186,
Japan (Tel:425-77-8960, Fax:425-77-3020)
FEATURES
    Location/Qualifiers
        source
            1..29392
            /organism="Mus musculus"
            /strain="129SVJ"
            /db_xref="taxon:10090"
            /chromosome="1"
            /map="1E3"
            /sex="female"
            /tissue_type="liver"
            2470..2474
            CAAT_signal
            2580..2585
            TATA_signal
            2878..2882
            CAAT_signal
            3176..3180
            CAAT_signal
            3251..3255
            CAAT_signal
            3350..3355
            5'UTR
            3389..3501
            /note="based on the putative transcription start"
            3389..3501
            /number=1
            exon
            3502..11067
            /number=1
            intron
            11124..27591
            /number=1
            gene
            /gene="pigr"
            join(11124..11166,18192..18536,21130..21783,23022..23354,
23908..24240,25133..25325,25686..25813,26183..26314,
26426..26484,27496..27591)
            /gene="pigr"
            /function="mediates secretion of IgA and IgM"
            /codon_start=1
            /product="polyimmunoglobulin receptor"
            /protein_id="BAA24431.1"
            /db_xref="GI:2804246"
            /translation="MRLYLFTLLVTVFSGVTKSPFGPQVWSVIEGDSVSICTCYPD
TSVNRHTRKYMCROGAGCMCTLISSNGYLSKEYSGRANLINFENNTFVINIQLTQ
DDTSYKCGLGTSNRGLSFDVLSVQVPELPSDTHVYTKDIGNRVIECFKRENAP
SKKSCKTKNSCELVIDSTKVNPSYIGRAKLFMKGTDLTFFVNIHSLTHNDAGLY
ICQAGEPSADKKNDVQLVLAPEPELKYKDLRSVTFECDLGRVANEAKYLCRMNKE
TCDVIINTLGRKDPDFEGRILITPKDDNGRFSVLITGLRKEDAGHYOCGAHSSGLPQE
GWPIOWLFVNEESTIPNRRSVKVTGGSVAIACPNPKESSLKYWCNWDGNG
HCPVLVGTQAOVQEEYEGRLALFDQPGNGTYTVILNLTTEDAGFYWCLTNGDSRWRT
TIELOVAETREPNELEVTPONATVILGETFTFVSCHYPCPKFYSQKYCKWKNKSGCHIL
FSDHGARGQSVSCDQSOLVSMTLNPKSDEGWYWCYKOGQTYGETTAIYIAVEE
RTRGSHVNPDTDANARKAVALVEEYVDSISSEKENKAIPNPGPFANEREIQNVDMQAO
ENRASGDAGSADGQSRSSSKVLFLSTLVLPLGLVLAAGIAVWVARVRRHKNVDRMSIS
SVRTDISMADFKNRDLGGNDNMGASPTQQTQVTIEGKDEIVTTTCTAEPESKAKR
SSKEADNAYSAFLQSSTIAAQVHDGPQEA"
            11124..11166
            /gene="pigr"
            /number=2
            intron
            11167..18191
            /gene="pigr"
            /number=2
            exon
            18192..18536
            /gene="pigr"
            /number=3
            intron
            18537..21129
            /gene="pigr"
            /number=3
            exon
            21130..21783
            /gene="pigr"
            /number=4
            intron
            21784..23021
            /gene="pigr"
            /number=4
            exon
            23022..23354
            /gene="pigr"
            /number=5
            intron
            23355..23907
            /gene="pigr"
```

JOURNAL Submitted (26-FEB-1997) to the DDBJ/EMBL/GenBank databases. Takashi
Sato, Yakult Central Institute for Microbiological Research,
Culture Collection Center; 1796 Yaho, Kunitachi-shi, Tokyo 186,
Japan (Tel:425-77-8960, Fax:425-77-3020)


```
170 GSTAACTCGGTCTCCATCATCGTCTACTACCCAGACACCTCTGTCAACGG 219
25 thIsValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||||:||||| ||| |||||:|: ||| ||||| |||:
220 GCACACCCGGAATACTGTCGCCAAGAGGCC...AACGGCTACTGG 266
42 lyThrValValSerThrThrAsnPhelIleAlaGlyThrGlyArg 58
||||:||||:||||:||||:||||:||||:||||:||||:||||:
267 CAACCTCATCTCTCAATGCTACTCTCGAAGGAGTATTCAAGCAGA 316
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
||||:||||| |||||:||||:||||:||||:||||:||||:||||:
317 GCAGGCTCATCACTCCAGAGATAGCACATTTGTTGATTAACATTGC 366
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMet 92
||||:||||| |||||:||||| ||| ||||| ||||:
367 ACATCTCACCCAGGAGACACTGGAGCTACAAGTGTGCTGGGT...A 413
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
||||:||||| |||||:||||| |||||:||||| |||||:
414 CCACTAACCCAGGCGCTGTTTTCGATGTCAGCCTGGAGTCT... 454
109 TyrGluProSerIrpGluGluGlnProMetProGluThrProLysIrpPh 125
||||:||||| |||||:||||| |||||: |||||: |||||:
455 ..... ACCAGGTTCCGAGTTCCTCCCAATGACAC 483
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerLysP 142
||||: |||||: |||||: |||||: |||||: |||||:
484 CCAATGTC.....TACACAAAGG 500
142 heValThrArgValThrProAlaGlnArgGlyLysValProVal 158
||||: |||||: |||||: |||||: |||||: |||||:
501 ACATAGCAGCAACTGTGACCATCGAATGCCGTTTCAAGAGGGAATGCT 550
159 HisHisSer_SerProThrThrGlnIleThrHisArgPro 171
||||: |||||: |||||: |||||: |||||: |||||:
551 CATAGCAGAATCCCTGTGTAGAAAGAGGAGGAGGCT 590
```

seq_name: gb_om:AB032195

```
seq_documentation_block: 2461 bp mRNA MAM 03-FEB-2000
LOCUS AB032195
DEFINITION Sus scrofa mRNA for poly-Ig receptor, complete cds.
ACCESSION AB032195
VERSION AB032195.2 GI:6863079
KEYWORDS poly-Ig receptor.
SOURCE Sus scrofa female cDNA to mRNA, clone:SC1.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 2461)
Sone,T. and Kumura,H.
TITLE Porcine mammary gland cDNA clone, similar to poly-Ig receptor
JOURNAL Published Only in DataBase (1999) In press
REFERENCE 2 (bases 1 to 2461)
Sone,T. and Kumura,H.
AUTHORS Direct Submission
TITLE Submitted (07-SEP-1999) to the DBJ/EMBL/GenBank databases. Teruo
JOURNAL Sone, Hokkaido University, Laboratory of Applied Microbiology,
Faculty of Agriculture; Kita-9, Nishi-9, Kita-ku, Sapporo, Hokkaido
060-8589, Japan (E-mail:sonet@chem.agr.hokudai.ac.jp,
Tel:81-11-706-2502, Fax:81-11-706-4961)
COMMENT On Feb 3, 2000 this sequence version replaced gi:5881588.
Sequence updated (01-Feb-2000).
```

```
FEATURES
source
    1..2461
        /organism="Sus scrofa"
        /db_xref="taxon:9823"
        /clone="SC1"
        /sex="female"
sig_peptide
    58..111
        /product="putative"
CDS
    58..2334
```

seq_name: gb_pr2:S43435S03

seq_documentation_block: 345 bp DNA PRI 08-MAY-1993

LOCUS S43435S03

DEFINITION transmembrane secretory component [human, leukocytes, Genomic, 345 nt, segment 3 of 11].

ACCESSION S43441

VERSION S43441.1 GI:255089

KEYWORDS

```
/codon_start=1
/product="poly-Ig receptor"
/protein_id="BAA84283.2"
/db_xref="GI:6863080"
/translation="MTRFFYACLLAIFPVVSMKSPIFGPDVSSVEGSSVIRCYPPA
TSVNRHSRYWCRIGAKGRCTLLISSEGIISKDYGRANLTNFPENGTFVMDIGHLTR
GDSGLYKGLGIGISSRGLSFDVLSYQSGPGQGVHVTADLGTSTVINCFFKSENAQ
KPKSVYKLGQIRLVIDSNGVNNNTNRAHLSIQGTNOLVFSFVNIQIRDAGIY
ICQADESSADLOVLKPEPELIGDLRGSVTFDCALGEMANVAKFLCOLKNGKTCN
VVTILGKRAQDFEGRLITLTKENSHFVHITGLKEDAGHYLCAGHPDGPKGMPV
QAWQLFINEDMIPRSSVVGGSVAVTCPPNPKETNSLKMYCRWEENENGRCPO
LIVESGLVNEQYEGRLAYEEPGNGTFTVILNLTNRDAGFYCLTNEDSDSRSTVEL
KIVEGEPNLKLPENVTAWGTETKLFSCHFPKCFYQYIKYCKWSNLTGRALPSQDEGQ
SOAFVNCDDKKSQIISLNLNPKRDEGWKCVDGLGHGETGVYVAVGKQKAGSGD
ARLANAAPAEDAIEPARPARENELVLDPSLIAEDRAKDGEGPAGSGVPAAPGSS
VGQGGSKRAVSTLPLVALMYLAVTIGVLBARHKRVDRISISYSTDLSMSDFES
SRDFGHDNAGATLDTOETSLGKDKRFTTTTNTMEDKEPRKRRKSSKEADKATFTT
LLQANSIIAATQNGPRE"
```

BASE COUNT 608 a 699 c 708 g 446 t

ORIGIN

alignment_scores:

Quality:	181.00	Length:	118
Ratio:	2.207	Gaps:	4
Percent Similarity:	69.492	Percent Identity:	33.898

alignment_block:

US-09-135-238B-2_COPY_18_253 x AB032195 ..

Align seg 1/1 to: AB032195 from: 1 to: 2461

```
2 IleLeuProGluValLysValGluGlyGluLeu..... 12
||||:||||| |||||:||||: |||||: |||||: |||||: |||||:
91 ATCTTCCCAGTGGTCTCCATGAAGAGTCCCATATTCGGTCCCCAGGATGT 140
13 .....GlyGlySerValThrIleLysCysProLeuPro.... 23
||||:||||| |||||:||||: |||||: |||||: |||||: |||||:
141 GAGCAGCGTGGAAAGCAGCTCGGTGTCCATCAGATGCTACTACCAGCCA 190
24 .....GluMetHisValArgIleTyrLeuCysArgGluMetAlaGly 37
::: ||| ||| ||| ||||| |||||: |||||: |||||: |||||:
191 CCTCGCTCAACCGGCACTTCTCGGAAGTACTGGTGCCGA...ATAGGAGCC 237
38 SerGlyThrCysGlyThrValValSerThrThrAsnPhelIleLysAlaG1 54
||||: ||| |||||:||||: |||||: |||||: |||||: |||||:
238 AAGGGCGCTGCACACACCTCATCTCTCGGAGGCGTACATCTCCAAGGA 287
54 uTyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheL 71
||||:||||| |||||:||||: |||||: |||||: |||||: |||||:
288 CTACAAGGCGCAGAGCCACCTCACCAACTTCCAGAGACGCGACCTTCG 337
71 euValGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCys 87
||||:||||| |||||:||||: |||||: |||||: |||||: |||||:
338 TGATGGACATTGGCCACCTGACCCCGGCGTGACTCTGGGCTCTACAAGTGT 387
88 GYAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAs 104
||| |||||:||||: |||||: |||||: |||||: |||||: |||||:
388 GGTCGGCATATGACAGC...CGAGGCTGTCTTTTGTGATGTGAGCCTGGA 434
104 nVal 105
|||
435 GGTC 438
```

```

SEGMENT      3 of 11
SOURCE       human leukocytes.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS      Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 345)
Molecular cloning and exon-intron mapping of the gene encoding
human transmembrane secretory component (the poly-Ig receptor)
JOURNAL      Eur. J. Immunol. 22 (9), 2309-2315 (1992)
MEDLINE      93387236
REMARK       GenBank staff at the National Library of Medicine created this
              entry [NCBI gisbq 113232] from the original journal article.
              This sequence comes from Fig. 2.
FEATURES     Location/Qualifiers
              source          1..345
              mat_peptide     /organism="Homo sapiens"
              /db_xref="taxon:9606"
BASE COUNT   78 a 102 c 97 g 68 t
ORIGIN
alignment_scores:
  Quality: 180.50      Length: 106
  Ratio: 2.314        Gaps: 4
  Percent Similarity: 73.585 Percent Identity: 38.679
alignment_block:
  US-09-135-238B-2_COPY_18_253 x S43435S03 ..
  Align seg 1/1 to: S43435S03 from: 1 to: 345
4 ProGluValLysValGluGlyGluLeuGlyGlySerValThrLeuLysCy 20
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30 CCCGAG...GAGGTGAATAGTGTGGAAGTAACTCAGTGCATCATCAGG 76
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
77 CTACTACCCACCCACCTCTGTCACCGGCACACCGGAGTACTGTGGCC 126
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
33 rGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
127 GCACGGAGGAGTAGAGTGCC...TGCAATAACCTCATCTCTCGGAGGCG 173
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
174 TAGGTCTCCAGCAATATGCAGCGCAGGCTAACTCACCACACTTCCCGGA 223
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
224 GACGGCACATTTGTGTGAACATTGCCAGCTGAGCCAGGATGACTCCG 273
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
83 lYValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
274 GCGGTACAAAGTGTGGCTGGGCATCAATAGC...CGAGGCTGTCTCTT 320
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
100 LysValThrLeuAsnVal 105
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
321 GATGTCAGCCTGGAGGTC 338
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name: gb_pat:A52094
seq_documentation_block:
  LOCUS      A52094      1936 bp      DNA
  DEFINITION Sequence 4 from Patent WO9618734.
  ACCESSION  A52094
  VERSION     A52094.1 GI:2304706
  KEYWORDS   unidentified.
  SOURCE      unidentified.
  ORGANISM   unclassified.
REFERENCE    1 (bases 1 to 1936)
AUTHORS      Van,O.J., Asseilbergs,F.A., Leist,C. and Gruetter,M.
TITLE       PRODUCTION OF RECOMBINANT SECRETORY COMPONENT
JOURNAL      CIBA GEIGY AG (CH)
REMARK       Other publication AU 4302796 960703.
FEATURES     Location/Qualifiers
              source          1..1936
              mat_peptide     /organism="unidentified"
              /db_xref="taxon:32644"

```

```

TITLE       PRODUCTION OF RECOMBINANT SECRETORY COMPONENT
JOURNAL      CIBA GEIGY AG (CH)
COMMENT      Other publication AU 4302796 960703.
FEATURES     Location/Qualifiers
              source          1..1936
              mat_peptide     /organism="unidentified"
              /db_xref="taxon:32644"
BASE COUNT   451 a 521 c 583 g 381 t
ORIGIN
alignment_scores:
  Quality: 180.50      Length: 106
  Ratio: 2.314        Gaps: 4
  Percent Similarity: 73.585 Percent Identity: 38.679
alignment_block:
  US-09-135-238B-2_COPY_18_253 x A52094 ..
  Align seg 1/1 to: A52094 from: 1 to: 1936
4 ProGluValLysValGluGlyGluLeuGlyGlySerValThrLeuLysCy 20
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
88 CCCGAG...GAGGTGAATAGTGTGGAAGTAACTCAGTGCATCATCAGG 134
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
135 CTACTACCCACCCACCTCTGTCACCGGCACACCGGAGTACTGTGGCC 184
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
33 rGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
185 GCGGAGGAGTAGAGTGCC...TGCAATAACCTCATCTCTCGGAGGCG 231
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 TAGGTCTCCAGCAATATGCAGCGCAGGCTAACTCACCACACTTCCCGGA 281
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
282 GACGGCACATTTGTGTGAACATTGCCAGCTGAGCCAGGATGACTCCG 331
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
83 lYValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
332 GCGGTACAAAGTGTGGCTGGGCATCAATAGC...CGAGGCTGTCTCTT 378
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
100 LysValThrLeuAsnVal 105
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
379 GATGTCAGCCTGGAGGTC 396
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name: gb_pat:A52093
seq_documentation_block:
  LOCUS      A52093      2031 bp      DNA
  DEFINITION Sequence 3 from Patent WO9618734.
  ACCESSION  A52093
  VERSION     A52093.1 GI:2304705
  KEYWORDS   unidentified.
  SOURCE      unidentified.
  ORGANISM   unclassified.
REFERENCE    1 (bases 1 to 2031)
AUTHORS      Van,O.J., Asseilbergs,F.A., Leist,C. and Gruetter,M.
TITLE       PRODUCTION OF RECOMBINANT SECRETORY COMPONENT
JOURNAL      CIBA GEIGY AG (CH)
REMARK       Other publication AU 4302796 960703.
FEATURES     Location/Qualifiers
              source          1..2031
              mat_peptide     /organism="unidentified"
              /db_xref="taxon:32644"

```


mat_peptide 111..2015
BASE COUNT 477 a 540 c 607 g 407 t
ORIGIN

alignment_scores:
Quality: 180.50 Length: 106
Ratio: 2.314 Gaps: 4
Percent Similarity: 73.585 Percent Identity: 38.679

alignment_block:
US-09-135-238B-2_COPY_18_253 x A52093 ..

Align seg 1/1 to: A52093 from: 1 to: 2031

```
4 ProGluValLysValGluGlyLeuGlyGlySerValThrIleLysCY 20
||||| :|||:||||| :|||:|||||:||||| ||
183 CCCGAG...GAGGTGAATAGTGTGGAAGTAACTCAGTGTCCATCACGTG 229
||| :|||:||||| :|||:|||||:||||| ||
20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33
||| :|||:||||| :|||:|||||:||||| ||
230 CTACTACCCACCCACTCTGTCAACCGGCACACCGGAAGTACTGGTGCC 279
||| :|||:||||| :|||:|||||:||||| ||
33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
||||| :|||:||||| :|||:|||||:||||| ||
280 GGCAGGGAGCTAGAGTGGC...TGCATAACCTCATCTCTCGGAGGGC 326
||| :|||:||||| :|||:|||||:||||| ||
50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66
||||| :|||:||||| :|||:|||||:||||| ||
327 TAGGTCTCCAGCAATATGACGAGCGAGGTAACTCACCACCTCCCGGA 376
||| :|||:||||| :|||:|||||:||||| ||
66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerg 83
||||| :|||:||||| :|||:|||||:||||| ||
377 GAACGGCACATTCTGGTGAACATTGCCAGCTGAGCCAGGATGACTCCG 426
||| :|||:||||| :|||:|||||:||||| ||
83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
||| :|||:||||| :|||:|||||:||||| ||
427 GCGCTACAAAGTGTGGCTGGGCATCAATAGC...CGAGGCTGTGCTTT 473
||| :|||:||||| :|||:|||||:||||| ||
100 LysValThrLeuAsnVal 105
||| :|||:||||| :|||:|||||:||||| ||
474 GATGTCAGCCTGGAGTGC 491
```

seq_name: gb_pat:A52091

seq_documentation_block:
LOCUS A52091 2405 bp DNA PAT 11-MAR-1997
DEFINITION Sequence 1 from Patent WO9618734.

ACCESSION A52091

VERSION A52091.1 GI:2304703

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS Van,O.J., Asselbergs,F.A., Leist,C. and Gruetter,M.

TITLE PRODUCTION OF RECOMBINANT SECRETORY COMPONENT

JOURNAL Patent: WO 9618734-A 1 20-JUN-1996;

COMMENT CIBA GEIGY AG (CH)

Other publication AU 4302796 960703.

FEATURES

1..2405 Location/Qualifiers

..

..

..

..

..

..

..

..

..

DDSGRYKCGLGINSGRLSPDVSLEVSQGGPGLNDTKVYTVDLGRTVTINCPFKTENAO
KRSLYKQIGLYPVLYIDSSGYNPNYTHGRLDIQGTGQLLESVIVNOLRLSDAGQY
LCQAGDDSNKKKNADLOVLPKPELVEDLRGVSFTHCALGFEVANAKFLCRQSSG
ENCDDVNTLGGKAPAFEPDGRILLNPQDKDSFVITGURKEDAGRYLCGAHSDGQLQ
EGSPIQAWQLFVNEESTIPRSPVTVKGVAGSVAVLCVPYNRKESKIKYWLWEGAON
GRCPLLVDSEGWKAQVEGRSLLEPEPNGTFTVILNQLTSDAGFYWCITNGDITLWR
TTVEIKIIEGEPNLKVPNGVTAVLGETLKVPCHPKFSSYKWKWNTGCOALPS
QDEGPKAFVNCDENSLVSILTNLNLTTRADEGHWGCVKOGHFYGETAAVYVAERK
AAGSRVSLAKADRAPEKVIDSGFREIENKATQDPRLFAEERKAVADTRQADGSRAS
VDGSSBEQGGSSRALVSTLVLPIGLVLAGAVAVGARHRKRNVDKRVSTRIDIS
MSDFNSREFGANDMGASITQETSLGGKEEFVATTESTTETKEPKKAKRSKEEAE
MAYKDFLLQSSTVAAEAQDGPQA"

BASE COUNT 579 a 648 c 718 g 460 t
ORIGIN

alignment_scores:
Quality: 180.50 Length: 106
Ratio: 2.314 Gaps: 4
Percent Similarity: 73.585 Percent Identity: 38.679

alignment_block:
US-09-135-238B-2_COPY_18_253 x A52091 ..

Align seg 1/1 to: A52091 from: 1 to: 2405

```
4 ProGluValLysValGluGlyLeuGlyGlySerValThrIleLysCY 20
||||| :|||:||||| :|||:|||||:||||| ||
183 CCCGAG...GAGGTGAATAGTGTGGAAGTAACTCAGTGTCCATCACGTG 229
||| :|||:||||| :|||:|||||:||||| ||
20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33
||| :|||:||||| :|||:|||||:||||| ||
230 CTACTACCCACCCACTCTGTCAACCGGCACACCGGAAGTACTGGTGCC 279
||| :|||:||||| :|||:|||||:||||| ||
33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
||||| :|||:||||| :|||:|||||:||||| ||
280 GGCAGGGAGCTAGAGTGGC...TGCATAACCTCATCTCTCGGAGGGC 326
||| :|||:||||| :|||:|||||:||||| ||
50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66
||||| :|||:||||| :|||:|||||:||||| ||
327 TAGGTCTCCAGCAATATGAGCGAGGCTAACCTCACCACCTCCCGGA 376
||| :|||:||||| :|||:|||||:||||| ||
66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerg 83
||||| :|||:||||| :|||:|||||:||||| ||
377 GAACGGCACATTCTGTGAACATTGCCAGCTGAGCCAGGATGACTCCG 426
||| :|||:||||| :|||:|||||:||||| ||
83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
||| :|||:||||| :|||:|||||:||||| ||
427 GCGCTACAAAGTGTGGCTGGGCATCAATAGC...CGAGGCTGTGCTTT 473
||| :|||:||||| :|||:|||||:||||| ||
100 LysValThrLeuAsnVal 105
||| :|||:||||| :|||:|||||:||||| ||
474 GATGTCAGCCTGGAGTGC 491
```


OM of: US-09-135-238B-2_COPY_18_253 to: EST:* out_format : pfs
Date: Sep 12, 2000 6:04 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame-p2n.model -DEV=xlp
-Q/cgn2.1/USPTO_spool/US09135238/runat_24082000_091323_17869/app_query.fasta_1.1282
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US09135238 @CGN1_1_2178 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-135-238B-2_COPY_18_253
Query length: 236
Database: EST:*
Database sequences: 5247842
Database length: -2090053206
Search time (sec): 1563.530000

score_list:

Sequence	Strd	Orig	2Score	EScore	Len	Documentation
gb_est42:AW402953	+	544.00	919.66	2.1e-42	312	! AW402953 UI-HF-BK0-abd-c-06-0-0
gb_est45:AW655210	+	451.00	756.39	2.7e-33	518	! AW655210 105863 MARC 1BOV Bos t
gb_est42:AW444691	-	443.50	746.13	9.9e-33	482	! AW444691 UI-H-BI3-ajw-g-12-0-UI
gb_est41:AW327067	+	415.00	699.20	4.1e-30	359	! AW327067 20516 MARC 1BOV Bos t
gb_est45:AW655218	+	384.00	643.17	5.4e-27	494	! AW655218 105873 MARC 1BOV Bos t
gb_est8:AA521933	+	364.50	607.52	5.2e-25	627	! AA521933 vhl61e2.r1 Soares_mamm
gb_est12:AA830944	-	305.00	509.27	1.5e-19	485	! AA830944 oc58b10.s1 NCI_CGAP_GC
gb_est12:AA491123	+	299.00	509.72	1.5e-19	479	! AA491123 ac46g04.r1 NCI_CGAP_GC
gb_est40:AW291041	-	292.50	486.49	2.9e-18	561	! AW291041 UI-H-BI2-agj-h-06-0-UI
gb_est25:AI1760373	-	242.00	398.24	2.4e-13	711	! AI1760373 wh87d11.x1 NCI_CGAP_CI
gb_est4:AA289657	+	220.50	367.44	1.2e-11	417	! AA289657 vb20c07.r1 Soares_mous
gb_est8:AA509857	+	209.50	356.67	4.9e-11	199	! AA509857 vhl5a10.r1 Soares_mamm
gb_est36:AW012833	+	180.00	301.29	5.9e-08	327	! AW012833 uo3d11.y1 Sugano_mous
gb_est36:AA174968	+	157.50	259.46	1.3e-05	460	! AA174968 ms78h03.r1 Soares_mous
gb_est15:AI021484	+	145.00	246.96	6.3e-05	220	! AI021484 ub08h03.r1 Soares_mamm
gb_est3:AA177660	+	145.00	246.31	6.9e-05	216	! AA177660 mt20d11.r1 Soares_mous
gb_est44:AW603666	+	127.00	204.60	0.0144	614	! AW603666 CM0-CN0039-280100-168
gb_gss15:AW033800	+	121.00	196.39	0.0403	501	! AW033800 RPCI-23-248K2.TV RPCI-
gb_est48:FI14851	+	120.50	201.00	0.0229	306	! FI14851 SSC24F11 Porcine small i
gb_est44:AW603303	+	120.50	193.57	0.0593	614	! AW603303 RCI-CN0012-170100-011
gb_est49:AW84070	+	117.00	191.41	0.0783	431	! AW84070 y94c09.r1 Soares retina
gb_est43:AW484927	+	107.50	175.71	0.5863	414	! AW484927 63153 MARC 3BOV Bos t
gb_est12:AA839051	+	105.50	170.54	1.14	489	! AA839051 vw47a02.r1 Soares_mamm
gb_est49:AW84076	+	105.00	181.73	1.00	411	! AW84076 y94c12.r1 Soares retina
gb_est17:AI177514	+	102.50	174.34	2.40	523	! AI177514 EST221146 Normalized r
gb_est47:AW603303	+	102.50	161.47	3.64	710	! AW603303 RCI-CN0012-170100-011
gb_est39:AW176541	+	101.50	168.34	1.51	318	! AW176541 MR0-CT0062-200899-002
gb_est40:AW258498	+	101.50	163.03	2.98	523	! AW258498 uq35h10.y1 NCI_CGAP_Ma
gb_est3:AA212711	+	100.50	162.34	3.26	476	! AA212711 mw81f05.r1 Soares_mous
gb_gss8:AA054923	+	100.50	162.03	3.39	490	! AA054923 RPCI-11-383F15.TV RPCI-
gb_gss4:AA0307015	+	100.00	153.53	10.08	1003	! AA0307015 HS-2046-BL-H02-77 CIT
gb_est44:AW603280	-	99.00	157.94	5.72	566	! AW603280 RCI-CN0012-170100-011
gb_est44:AW603762	-	96.00	155.02	8.33	462	! AW603762 QV0-ST0247-200100-092
gb_est46:AW710271	+	96.00	154.48	8.93	486	! AW710271 e2e08ne.r1 Neurospora
gb_gss11:AW751079	+	95.50	142.87	39.58	1332	! AW751079 HS-5576-BL-G04.SP6.RB
gb_est42:AW407899	+	95.00	142.94	8.41	397	! AW407899 UI-HF-BM0-adk-f-09-0-UI
gb_est13:AW027770	+	94.50	146.10	26.13	839	! AW027770 ov99e09.x1 Soares_test
gb_est24:AW402453	+	93.50	149.70	16.48	511	! AW402453 ui-HF-BK0-aan-f-12-0-UI
gb_est24:AW1746782	+	93.00	152.44	11.59	365	! AW1746782 ui07d03.y1 Sugano_mous
gb_est44:AW604875	+	93.00	151.65	12.83	393	! AW604875 QV1-CT0366-280100-055
gb_est46:AW726179	+	92.50	141.70	45.96	922	! AW726179 GA_Ea0020006 Gossypiu

gb_est17:AI153762 + 92.00 147.98 20.55 473 ! AI153762 ud48g01.r1 Soares_m
gb_est24:AI685262 - 92.00 145.93 26.72 573 ! AI685262 wa75e11.x1 Soares_N
gb_est5:AA306437 + 91.50 146.52 24.79 501 ! AA306437 EST177393 Jurkat T-
gb_est54:W42194 - 91.50 138.06 73.30 1106 ! W42194 mc69a10.r1 Soares mo
seq_name: gb_est42:AW402953
seq_documentation_block: 312 bp mRNA EST 16-FEB-2000
LOCUS AW402953
DEFINITION UI-HF-BK0-abd-c-06-0-UI.r1 NIH_MGC_36 Homo sapiens CDNA clone
IMAGE:3055859 5', mRNA sequence.
ACCESSION AW402953
VERSION AW402953.1 GI:6921743
KEYWORDS EST..
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 312)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138550.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnlni.gov/bbrp/image/image.html
Seq primer: M13 Forward.

FEATURES

Location/Qualifiers
1..312
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3055859"
/clone_lib="NIH_MGC_36"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LRI)"
/note="Vector: p7T3-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(0.5-1.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 97 a 74 c 87 g 54 t

alignment_scores:

Quality: 544.00 Length: 101
Ratio: 5.386 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-135-238B-2_COPY_18_253 x AW402953 ..
Align seg 1/1 to: AW402953 from: 1 to: 312
24 GLuMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyth 40
|||||
10 GAAATGTCATGTGAGGATATATCTGTCCGGAGATGCTGGATCTGGAAC 59
40 rCysGlyThrValValSerThrThrAsnPheLeuLeuAlaGluTyrLysG 57
|||||
60 ATGTGGTACCTGGTGTCCACCACTTCATCAAGGAGAGATCAACAGG 109
57 lYArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 73

110 GCGAGTTACTCTGAACAATATACCCACGAAGAATCTGTTCTAGTGGAG 159
74 ValThrGlnLeuThrGluSerAspSerGlyValTyAlaCysGlyAlaG1 90
160 GTAACACAGCTGACAGAAAGTGACACCGAGTCTATGCTGCGGAGCGG 209
90 yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 107
210 CATGAACACAGACCGGGAAGACCCAGAAAGTCACCCCTGAATGTCCACA 259
107 exGluTyGluProSerTrpGluGluGlnProMetProGluThrProLys 123
260 GTGATAGACCCATCATCGGGAAGAGACCAATGCTGAGACTCCAAA 309
124 Ttp 124
310 TGG 312
seq_name: gb_est45:AW655210
seq documentation_block:
LOCUS AW655210 518 bp mRNA EST 05-APR-2000
DEFINITION 105863 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW655210
VERSION AW655210.1 GI:7421036
KEYWORDS EST.
SOURCE Bos taurus.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 518)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
Keele,J.W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
Unpublished (2000)
On Mar 8, 1999 this sequence version replaced gi:4388119.
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCGCCAGTCACGACG
Plate: 86 row: K column: 10
Seq primer: ATTTAGGTGACACTATAG.
FEATURES
source
1..518
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10b"
/note="Vector: pCMV SPORT6; Site 1: XbaI; Site 2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 124 a 133 c 157 g 104 t
ORIGIN
alignment_scores:
Quality: 451.00 Length: 119
Ratio: 4.295 Gaps: 0
Percent Similarity: 88.235 Percent Identity: 67.227

alignment_block:
US-09-135-238b-2_copy_18_253 x AW655210 ...
Align seg 1/1 to: AW655210 from: 1 to: 518
1 ArgTleLeuProGluValLysValGluGlyGluLeuGlyGlySerValTh 17
161 AAGTCTCTCCAGAAAGTAAAGATGAAGGAATGCTGGGAGGATCCATTAC 210
17 rLleLysCysProLeuProGluMetHisValArgTleTyLeuCysArgG 34
211 CATGAGTGCACCTCTCTGAAAGCATGTGAGGATATATCTGTGCGGGA 260
34 luMetaLagLysSerGlyThrCysGlyThrValValSerThrThrAsnPh 50
261 CCATTGACGAGTCTGGAAGATGCACCGCTGCTGCCAGCAACAATAC 310
51 lLleLysAlaGluTyLysGlyArgValThrLeuLysGlnTyProArgLy 67
311 GTCAGAGAGGATTCACGCCGAGTCACGCTGGAGCAGTGTCCGGACAG 360
67 sasLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84
361 GAATCTGTTCTCTGTGTGATGACAGAGCTGACCAAGAATGACAGCGGA 410
84 alTyAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
411 TCTATGCTGTGGGTGGCGGACACAGACCCGAGGAGCAGACCCAGCAG 460
101 ValThrLeuAsnValHisSerGluTyGluProSerTrpGluGluGlnPr 117
461 ATCACCCTGACTGTTACAGCGCTTACAGAGCCATCTCTGGGAAGAGGCC 510
117 oMetPro 119
511 GATGCT 517
seq_name: gb_est42:AW444691
seq documentation_block:
LOCUS AW444691 482 bp mRNA EST 17-FEB-2000
DEFINITION UI-H-BI3-ajw-g-12-0-UI-s1 NCI CGAP_Sub5 Homo sapiens cDNA clone
IMAGE:2733502 3', mRNA sequence.
ACCESSION AW444691
VERSION AW444691.1 GI:6986453
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 482)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html The following repetitive
elements were found in this cDNA sequence: 1-30,
>POLY_A\$Simple_repeat
Seq primer: M13 Forward
POLYA=Yes.
FEATURES
source
1..482
/organism="Homo sapiens"
/db_xref="taxon:9606"

```
/clone="IMAGE:2733502"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub5
is a subtracted library derived from NCI_CGAP_Sub4. The
NCI_CGAP_Sub5 library had 3 million recombinants. A
single-stranded DNA preparation of NCI_CGAP_Sub4 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE Clonoids
1322376-1323911, 1456008-1456775, 1500552-1502855);
NCI_CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
(IMAGE Clonoids 1323912-1325831, 1471368-1472903
1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582,
3851-3854 (IMAGE Clonoids
1414920-1417991, 1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM
3164-3167, 3716-3720, 3733-3735 (IMAGE Clonoids
1257096-1258631, 1469064-1470983, 1475592-1476743);
NCI_CGAP_Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068
(IMAGE Clonoids 985608-986759, 1101192-1101959,
1217928-1220615); NCI_CGAP_Co10 pool 1 LLAM 2644-2653,
2871-2872 (IMAGE Clonoids
1057416-1061255, 1144584-1145351). (10% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (10% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (10% of the
driver population), plus a pool of 5,472 clones from
NCI_CGAP_Sub4 (IMAGE Clonoids 2723592-2728969) (70% of the
driver population). Subtraction was performed as
previously described [Bonaldo, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806. ;
TAG_LIB=NCI_CGAP_Kid5; TAG_TISSUE=Kidney; TAG_SEQ=ATTC"
BASE COUNT 103 a 107 c 111 g 161 t
ORIGIN

alignment_scores:
  Quality: 444.50      Length: 117
  Ratio: 4.401        Gaps: 3
  Percent Similarity: 86.325      Percent Identity: 80.342

alignment_block:
US-09-135-238B-2_COPY_18_253 x AW444691/rev ..
Align seg 1/1 to reverse of: AW444691 from: 1 to: 482

24  GluMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyTh 40
|||||
481  GAAATGCATGTGAGGATATCTGTCCGGGAGATGGCTGGATCTGGAAC 432
|||||
40  rCysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysG 57
|||||
431  ATGTGTACGTGGTATCCACCACCACTTCATCAAGGCAGATACAAAG 382
|||||
57  lyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 73
|||||
381  GCCGAGTTACTCTGAAGCAATACCCAGCAAGATCTGTTCTAGTGGAG 332
|||||
74  ValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGl 90
|||||
331  GTAACACAGCTGACAAAGTGTACACGGAGTCTATGCTCGGAGCGGG 282
|||||
90  yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 107
|||||
281  CATGAACACAGACCGGGGAAGAACCCAGAAAGTCACCTGAATGTCCACA 232
|||||
107  er...GluTyrGluProSerTrpGluGluGlnProMetProGluThrPro 122
|||
231  GTGGTAGGTTCCCTCCGCTGATTGGAGCTCAG.....ACTACCCAG 191
```

```
123  LysTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSe 139
|||||
190  AAAATGTTTCACTTGGGAAGATTG...GAACAGCCATCTAACTGTAACAG 144
139 r 139
143 T 143

seq_name: gb_est41:AW327067

seq_documentation_block:
LOCUS AW327067 359 bp mRNA EST 27-JAN-2000
DEFINITION 20516 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW327067
VERSION AW327067.1 GI:6762988
KEYWORDS EST.
SOURCE Bos taurus.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 359)
AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
Keefe,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT On Jan 6, 2000 this sequence version replaced gi:6676592.
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCGCAGTCACGAGC
Plate: 13 row: K column: 5
Seq primer: ATTTAGGTGACACATATAG.
FEATURES
    source
        location/Qualifiers
            1..359
                /organism="Bos taurus"
                /db_xref="taxon:9913"
                /clone_lib="MARC 1BOV"
                /tissue_type="pooled"
                /lab_host="DH10B"
                /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                Library made from pooled tissue from lymph node, ovary,
                fat, hypothalamus, and pituitary."
BASE COUNT 90 a 111 c 87 g 71 t
ORIGIN

alignment_scores:
  Quality: 415.00      Length: 119
  Ratio: 4.192        Gaps: 1
  Percent Similarity: 83.193      Percent Identity: 64.706

alignment_block:
US-09-135-238B-2_COPY_18_253 x AW327067 ..
Align seg 1/1 to: AW327067 from: 1 to: 359

59  ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||||
2  GTCACGCTGGACGAGTGTCCGGACAGGAATCTGTTCTCTGTGGTGATGAC 51
|||||
75  rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
```

```
|||||.....|||||.....|||||.....|||||.....|||||.....
52 AGAGCTGACCAAGATGACAGCGGATCTATGCTGTGGGTGGCGGGA 101
|||||.....|||||.....|||||.....|||||.....|||||.....
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
|||||.....|||||.....|||||.....|||||.....|||||.....
102 ACACAGCCGAGGAGAACCCAGCAGATCACCTGACTGTTCACAGCGTT 151
|||||.....|||||.....|||||.....|||||.....|||||.....
109 TyGluProSerTrpGluGluGlnProMetProThrProLysTrpPh 125
|||||.....|||||.....|||||.....|||||.....|||||.....
152 TAGAGCCATCTGGGAGGAGGCGGATGCTGAGCTCCAGCATGGTT 201
|||||.....|||||.....|||||.....|||||.....|||||.....
125 e.....HisLeuProTyThrLeuPheGlnMetProAlaIat 136
|||||.....|||||.....|||||.....|||||.....|||||.....
202 TAATAGATTCTACAAATGCATCTGCTTGGTTCAGATGCTGCAC 251
|||||.....|||||.....|||||.....|||||.....|||||.....
136 yrlaSerSerLysPheValThrArgValThrProAlaGlnArg 152
|||||.....|||||.....|||||.....|||||.....|||||.....
252 ATGCCAGTTCTTTAAATTCACGTCAAGTAACACACACAGCTCAAGG 301
|||||.....|||||.....|||||.....|||||.....|||||.....
153 GlyLysValProValHisHisSerProThrThrGlnIleThrH1 169
|||||.....|||||.....|||||.....|||||.....|||||.....
302 ATCAAGTCTCTCGACACACAGGCTCCCCCAACCCCTCAGTCACCCA 351
|||||.....|||||.....|||||.....|||||.....|||||.....
169 sArgPro 171
|||||.....|||||.....|||||.....|||||.....|||||.....
352 CC8CCCC 358

seq_name: gb_est45:AW655218

seq_documentation_block:
LOCUS AW655218 494 bp mRNA EST 05-APR-2000
DEFINITION 105873 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.
ACCESSION AW655218
VERSION AW655218.1 GI:7421044
KEYWORDS EST.
SOURCE Bos taurus.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 494)
Smith,R.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
Keele,J.W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
Unpublished (2000)
Contact: Smith TPL
On Mar 8, 1999 this sequence version replaced gi:4388128.
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt.trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 86 row: M column: 8
Seq primer: ATTTAGGTGACACTAG.
FEATURES
Source
1..494
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="Dhl08"
/note="Vector; PCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 118 a 125 c 150 g 101 t
```

```
ORIGIN
alignment_scores:
Quality: 384.00 Length: 111
Ratio: 4.042 Gaps: 0
Percent Similarity: 85.586 Percent Identity: 63.063

alignment_block:
US-09-135-238B-2_COPY_18_253 x AW655218 ..
Align seg 1/1 to: AW655218 from: 1 to: 494
1 ArgIleLeuProGluValLysValGluGlyGluLeuGlyGlySerValTh 17
162 AAGGTCTTCCAGAGTAAGATGAAGAAATGCTGGGAGGATCCATTAC 211
|||||.....|||||.....|||||.....|||||.....|||||.....
17 rIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 34
|||||.....|||||.....|||||.....|||||.....|||||.....
212 CATCGAGTGCCCACTCTCTGAAACGCATGTGAGGATATATCTGTGCGGA 261
|||||.....|||||.....|||||.....|||||.....|||||.....
34 LuMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
|||||.....|||||.....|||||.....|||||.....|||||.....
262 CCATGTGACGAGCTGGAAGATGCACCACTGTGTCCAGCAACAATAC 311
|||||.....|||||.....|||||.....|||||.....|||||.....
51 IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgL 67
|||||.....|||||.....|||||.....|||||.....|||||.....
312 GTAAGAGAGGAATTCAAGCAGGAGCAGCTGGAGCAGTGTCCGGACAG 361
|||||.....|||||.....|||||.....|||||.....|||||.....
67 sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84
|||||.....|||||.....|||||.....|||||.....|||||.....
362 GAATCTGTCTTGTGTGATGACAGAGCTGACCAAGAAATGACAGCGGA 411
|||||.....|||||.....|||||.....|||||.....|||||.....
84 alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
|||||.....|||||.....|||||.....|||||.....|||||.....
412 TCTATGCTGTGGGTGGCGGGAACACAGACGAGGAGCAAGACCATCAG 461
|||||.....|||||.....|||||.....|||||.....|||||.....
101 ValThrLeuAsnValHisSerGluTyrGluPro 111
|||||.....|||||.....|||||.....|||||.....|||||.....
462 ATCACCCTGACTGTTACAGCGTTTACTAGCCA 494

seq_name: gb_est8:AA521993

seq_documentation_block:
LOCUS AA521993 627 bp mRNA EST 17-JUL-1997
DEFINITION vhl61e12.r1 Soaresmammary.gland_NbMMG Mus musculus CDNA clone
IMAGE:891502 5', mRNA sequence.
ACCESSION AA521993
VERSION AA521993.1 GI:2262738
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 627)
Marra,M., Hillier,D., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On Jan 14, 1998 this sequence version replaced gi:1797443.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:519462
```


162 SerProThrThrGlnIleThrHisArgProArgValSerArgAlaSerSe 178
 |||||
 485 TCCGCCACCCACCAATACCCACCGCCCTCGAGTGTCAGAGCATCTTC 436
 |||||
 178 rValAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerL 195
 |||||
 435 AGTAGCAGGTGACAAAGCCCGAACCTTCTGCCATCCACTACAGCCTCAA 386
 |||||
 195 ystIleSerAlaLeuGluGlyLeuLeuLysProGlnThrProSerTyrAsn 211
 |||||
 385 AAATCTCAGCTCTGGAGGGCTGCTCAAGCCCCAGAGCCAGCTACAAAC 336
 |||||
 212 HisHisThrArgLeuHisArgGlnArg 220
 |||||
 335 CACCACACAGGCTGCACAGGCGAGG 309

seq_name: gb_est8:AA491123

seq_documentation_block: 179 bp mRNA EST 18-AUG-1997
 LOCUS AA491123 aa46g04.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:824022 5',
 mRNA sequence.

ACCESSION AA491123
 VERSION AA491123.1 GI:2220296
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 179)

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

TITLE Unpublished (1997)

JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 CGAP Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -28m13 rev1 Et from Amersham
 High quality sequence stop: 116.

FEATURES

source

1. .179
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:824022"
 /clone_lib="NCI_CGAP_GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, IgD-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
 primed with a Not I - oligo(dT) primer
 [5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3/
]. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 46 a 69 c 34 g 30 t

BASE COUNT
 ORIGIN

alignment_scores:

Quality: 299.00 Length: 59
 Ratio: 5.068 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-135-238B-2_COPY_18_253 x AA491123 ..

Align seg 1/1 to: AA491123 from: 1 to: 179

146 ValThrThrProAlaGlnArgGlyLysValProProValHisHisSerSe 162
 |||||

3 GTTACCACACAGCTCAAAGGGGCAAGTCCCTCCAGTTCACCACCTCCTC 52
 |||||

162 rProThrThrGlnIleThrHisArgProArgValSerArgAlaSerSeV 179
 |||||

53 CCCACCACCCAAATCACCACCGACCTCGAGTGTCAGAGCATCTTCAG 102
 |||||

179 alAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLys 195
 |||||

103 TAGCAGGTGACAAAGCCCGAACCTTCTGCCATCCACTACAGCCTCAAAA 152
 |||||

196 IleSerAlaLeuGluGlyLeuLeuLys 204
 |||||

153 ATCTCAGCTCTGGAGGGGCTGCTCAAG 179

seq_name: gb_est40:AW291041

seq_documentation_block: 561 bp mRNA EST 16-JAN-2000

LOCUS AW291041 UT-H-B12-agi-h-06-0-01.s1 NCI_CGAP_Sub4 Homo sapiens cDNA clone

DEFINITION IMAGE:2724707 3', mRNA sequence.

ACCESSION AW291041

VERSION AW291041.1 GI:6697677

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 561)

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

TITLE Unpublished (1997)

JOURNAL On May 18, 1998 this sequence version replaced gi:3137211.

COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. cDNA library preparation: M.B. Soares Lab Clone distribution:
 NCI-CGAP clone distribution information can be found through the
 I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward

POLYA-Yes.

Location/Qualifiers

1. .561

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2724707"

/clone_lib="NCI_CGAP_Sub4"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker: Site_1: Not I; Site_2: Eco RI; The

NCI_CGAP_Sub4 library is a subtracted library derived from

the NCI_CGAP_Sub2 library which is a subtracted library

derived from the NCI_CGAP_Sub1 library, which is a

subtracted library derived from BI. BI constitutes a

mixture of 21 normalized or subtracted

libraries: NCI_CGAP_C04, NCI_CGAP_P-22, NCI_CGAP_P-28,

NCI_CGAP_C010, NCI_CGAP_C016, NCI_CGAP_K1d5,

NCI_CGAP_Kid12, NCI_CGAP_Kid3, NCI_CGAP_Kid11,
 NCI_CGAP_Lym2, NCI_CGAP_Br2, NCI_CGAP_G08, NCI_CGAP_CLL1,
 NCI_CGAP_Le12, NCI_CGAP_Brn23, NCI_CGAP_Lu5,
 NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
 NCI_CGAP_Brn25. These 21 libraries were pooled and a
 single-stranded DNA preparation of the resulting mixture
 was used as a tracer in a subtractive hybridization with
 a driver whose composition is detailed below:
 NCI_CGAP_Kid3 pool 1 : LLAM 3334-3337, 3682-3683,
 3798-3803 (IMAGE CloneIDs 1322376-1323911),
 1456008-1456775, 1500552-1502855) NCI_CGAP_Kid5 pool 1 :
 LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE CloneIDs
 1323912-1323831, 1471368-1472903, 1492104-1493255)
 NCI_CGAP_Lu5 pool 1 : LLAM 3575-3582, 3851-3854 (IMAGE
 CloneIDs 1414920-1417991, 1520904-1522439) NCI_CGAP_GC4
 pool 1 : LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
 CloneIDs 1257096-1258631, 1469064-1470983),
 1475592-1476743) NCI_CGAP_Pr22 pool 1 : LLAM 2457-2459,
 2758-2759, 3062-3068 (IMAGE CloneIDs 985608-986759,
 1101192-1101959, 1217928-1220615) NCI_CGAP_Co10 pool 1 :
 LLAM 2644-2653, 2871-2872 (IMAGE CloneIDs 1057416-1061255,
 1144584-1145351) Subtraction was performed as previously
 described [Bonaldo, Lennon & Soares (1996) : Normalization
 and Subtraction: Two Approaches To Facilitate Gene
 Discovery. Genome Research 6, 791-806.];
 TAG_Lib=NCI_CGAP_Co10; TAG_TISSUE=colon; TAG_SEQ=AAACG"

BASE COUNT 121 a 125 c 120 g 194 t 1 others
 ORIGIN

alignment_scores:
 Quality: 292.50 Length: 91
 Ratio: 3.953 Gaps: 3
 Percent Similarity: 81.319 Percent Identity: 72.527

alignment_block:

US-09-135-238B-2_COPY_18_253 x AW291041/rev ..

Align seg 1/1 to reverse of: AW291041 from: 1 to: 561

50 PheLeLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66
 |||||
 561 TTCAATCAAGGCGAGTATCAAGCGCGAGTTACTCTGAAGCAATACCCAGC 512
 66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83
 |||||
 511 CAAGATCTGTCTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCG 462
 83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
 |||||
 461 GAGTCTATGCTGCGGAGCGGCATGAACACAGACAGCCGNGAAGACCCAG 412
 100 LysValThrLeuAsnValHisSer... GluTyrGluProSerTrpGluG1 115
 |||||
 411 AAAGTCACCTGAAATGTCACAGTGGTAGGTCTCTCGCTGATTGGAGGC 362
 115 uGlnProMetProGluThrProLysTrpPheHisLeuProTyrLeuPheG 132
 |||||
 361 TCAG.....ACTACCCAGAAAATGTTTCACCTCGGGAAGATTG...G 324
 132 InMetProAlaTyrAlaSerSer 139
 :::::|||||
 323 AACAGCCATCTAACTGTAACTAGT 301

seq_name: gb_est25:AI760373

seq_documentation_block:

LOCUS AI760373 711 bp mRNA EST 20-DEC-1999
 DEFINITION wh87d11.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2387733 3'
 similar to TR:060667 060667 ANTI-FAS-INDUCED APOPTOSIS. ;, mRNA
 sequence.
 accession AI760373
 version AI760373.1 GI:5176040

KEYWORDS
 SOURCE

ORGANISM

REFERENCE
 AUTHORS
 TITLE

JOURNAL

COMMENT

On Oct 30 1998 this sequence version replaced gi:3812483.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,

M.D., Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1449 Std Error: 0.00

Seq primer: -400P from Gibco

High quality sequence stop: 472.

Location/Qualifiers

1. /711

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2387733"

/clone.lib="NCI_CGAP_CLL1"

/tissue_type="B-cell, chronic lymphocytic leukemia"

/lab_host="DH10B"

/note="Vector: pT730-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGAGCGCGCATGCTTTTTTTTTTTTTTTTTT

T 3']; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pT73 vector.

Library is normalized, and was constructed by Bento

Soares and M.Fatima Bonaldo."

BASE COUNT 127 a 195 c 239 g 148 t 2 others

ORIGIN

alignment_scores:

Quality: 242.00 Length: 65

Ratio: 4.246 Gaps: 1

Percent Similarity: 87.692 Percent Identity: 78.462

alignment_block:

US-09-135-238B-2_COPY_18_253 x AI760373/rev ..

Align seg 1/1 to reverse of: AI760373 from: 1 to: 711

173 ValSerArgAlaSerSerValAlaGlyAspLysProArgThrPheLeuPr 189

|||||

711 GTGTCAAGCATTCTCAAGTAGCAGGTGACAAAGCNCGCCACCTCTTTGCC 662

189 oSerThrAlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProG 206

|||||

661 ATCCACTACAGCTCAAAAATCTCAGCTCTGGAGGGCTGCTCAAGCCCC 612

206 InThrPro.SerTyrAsnHisHisThrArgLeuHisArgGlnArgAlaLe 222

|||||

611 AGAGCCAGGCTACACACACACACACAGCGGTGCACAGGCAGAGACACT 562

222 uAspTyrGlySerGlnSerGlyArgGluGlyGlnGlyPheHis 236

|||||

561 GGACTATGCTCAGCTCTGGAGGGAAGCCAGGATTTCAC 519

seq_name: gb_est4:AA289667

```
seq_documentation_block: 417 bp mRNA EST 14-APR-1997
LOCUS AA289667 v20c07.r1 Soares mouse 3nbms Mus musculus cDNA clone IMAGE:749484
DEFINITION 5', mRNA sequence.
ACCESSION AA289667
VERSION AA289667.1 GI:1937343
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 417)
AUTHORS Maria.M., Hillier.L., Allen.M., Bowles.M., Dietrich.N., Dubuque.T.,
Geisel.S., Kucaba.T., Lacy.M., Le.M., Martin.J., Morris.M.,
Schellenberg.K., Steptoe.M., Tan.F., Underwood.K., Moore.B.,
Theising.B., Wylie.T., Lennon.G., Soares.B., Wilson.R. and
Waterston.R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:458468
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 405.
FEATURES
source
1..417
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="Soares mouse 3NBMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I oligo(dT) primer [5'
TGTACCAATCTGAGTGGAGCGCCGCGTGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."
BASE COUNT 109 a 140 c 91 g 77 t
ORIGIN
alignment_scores:
Quality: 220.50 Length: 92
Ratio: 3.106 Gaps: 1
Percent Similarity: 77.174 Percent Identity: 47.826
alignment_block:
US-09-135-238B-2_COPY_18_253 x AA289667 ..
Align seg 1/1 to: AA289667 from: 1 to: 417
145 ArgValThrProAlaGlnArgGlyLysValProValHisHis 161
4 AAAGTTACCGCGCAACTCCAAAGACTGAGGCCCTCCGGTTCCACGACC 53
161 rSerProThrGlnIleThrHisArgProArgValSerArgAlaSer 178
|||||
|||||
```

```
54 CTCGAGCATCTTCAGTACCCAAACATCCAGAGTTCACAGAGCATTTT 103
178 erValAlaGlyAspIleProArgThrPheLeuProSerThrThrAlaSer 194
|||||
104 CTGTGTCCGTACCTACGAGTCCCGGCGTCTCTGCAGCAACACAGCCTCA 153
195 LysIleSerAlaLeuGluGlyLeuLeuLysProGlnThrProSerTyra 211
|||||
154 AAGACTTCCACTCAGCAAGCA...ATCAGGCCCTAGAGCCGCTACAG 200
211 nHisThrArgLeuHisArgGlnArgAlaLeuAspTyrglySerGlns 228
|||||
201 CCACACACACAGAGTCTCATGAGCAAGAGCAGCCACCATGCCCACT 250
228 erGlyArgGluGlyGlnGlyPheHis 236
|||||
251 ATGGGAGAGAGACCGAGGCTTCAC 276
seq_name: gb_est8:AA509857
seq_documentation_block: 199 bp mRNA EST 08-JUL-1997
LOCUS AA509857 vhs1a10.r1 Soares mammary_gland_NbMMG Mus musculus cDNA clone
DEFINITION IMAGE:890490 5' similar to SW:PIGR_HUMAN P01833
POLYMERIC-IMMUNOGLOBULIN RECEPTOR ;, mRNA sequence.
ACCESSION AA509857
VERSION AA509857.1 GI:2247711
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 199)
AUTHORS Marra.M., Hillier.L., Allen.M., Bowles.M., Dietrich.N., Dubuque.T.,
Geisel.S., Kucaba.T., Lacy.M., Le.M., Martin.J., Morris.M.,
Schellenberg.K., Steptoe.M., Tan.F., Underwood.K., Moore.B.,
Theising.B., Wylie.T., Lennon.G., Soares.B., Wilson.R. and
Waterston.R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:518450
Possible reversed clone: similarity on wrong strand
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 1.
FEATURES
source
1..199
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="Soares mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Organ: mammary gland; Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
TGTACCAATCTGAGTGGAGCGCCGCGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
```

constructed and normalized by Bento Soares and M. Fatima
Bonaldo "

BASE COUNT	55 a	40 c	58 g	46 t
------------	------	------	------	------

alignment_scores:		
Quality:	209.50	Length: 67
Ratio:	3.741	Gaps: 1
Percent Similarity:	83.582	Percent Identity: 62.687

```
alignment_block:
US-09-135-238B-2_COPY_18_253 x AA509857
```

Align seg 1/1 to: AA509857 from: 1 to: 199

27 ValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysGlyTh 43
|||||:::|||||:::|||||:::|||||:::|||||
2 GTAAGGATGTATCTGTGCGCAGATGCCCAACCTGGGATATGCTCCAC 51

43 rValValSerThrThrAsnPhelLeLysAlaGluTrpLysGlyArgValt 60
|||||
52 TETGGTGTCCACACC...TTTGTCAGAAGGAATATGAAGGGCAGTCA 98

60 hrLeuLysGlnTyrProArgLysAsnIeuPheLeuValGluValThrGln 76
||||| :||:::||:::||:::||:::||:::||:::||:::||:::||
99 CCCTGAGCCATGCTTGATAGAAGCTATTCTACTGGAGATGACACAG 148

77 LeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetAsnTh 93
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
149 CTACGGATAATGACCATGGAATCTATGCCCTGTGGTGCCATGAAGAC 198

93 r 93
—
199 A 199

seq_name: gb_est36:AW012833

seq_documentation_block:	327 bp	mrna	EST	10-SEP-1999
LOCUS	AW012833			
DEFINITION	uo03dl1.yl Sugano mouse liver mlia Mus musculus cDNA clone			
IMAGE:	2582325	5'	similar to gb:U06431 Mus musculus B6/CBAF1J	
polymERIC immunoglobulin receptor mRNA, (MOUSE);, mRNA sequence.				

ACCESSION AWO12833
VERSION AWO12833.1 GI:5861611
KEYWORDS EST.

SOURCE	ORGANISM
house mouse.	<i>Mus musculus</i>
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 327)
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Willson, R.

TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT On Jul 9, 1999 this sequence version replaced gi:5433821.
Contact: Maria M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Fax: 314 280 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:1021913

Seq primer: custom primer used
High quality sequence stop: 272.
Location/Qualifiers
1..327
/organism="Mus musculus"

```

/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2582325"
/clone.lib="Sugano mouse liver mlia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="organ: liver; Vector: pME18S-FL3; Site.1: DraIII
(CACATGTGTG); site.2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCCTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCGTCTCTAAAGCTGCG and 3' end
primer CGACCTCAGCTCGACACA."
97 a 98 c 76 g 66 t

```

```
alignment_scores:
  Quality: 180.00
  Ratio: 2.759
  Percent Similarity: 73.864
  Gaps: 3
  Percent Identity: 40.909
  Length: 88
```

```
alignment_block:
US-09-135-238B-2_COPY_18_253 x AW012833
```

Align seg 1/1 to: AW012833 from: 1 to: 327

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25

60 GCGACTCTGTTCCATCAGTGCTACTACCCAGACACCTCTGTCAACCG 109

25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
 ||||| ||| |||||: ||| ||||| |||
 110 GCACACCCGGAAATACTGGTCCGACAAAGGACC...ACGGCATGTGCA 156

42 lyThrValValSerThrThrAsnPheIleLysAlaGluTyLysGlyArg 58
 |||:::||||:::||||:::||||:::||||:::||||:::||||:::||||
 1157 CAAGCTCATCTCTTCCAAATGGCTACCTCTCTCCAGGAGTATTCAGGCAGA 206

59 ValThrIeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
:::||||| :::::||::||::||| |||:::||||:
207 GCCAACCTCATCAACTTCCCCAGACAAACACATTGTGATTAACTGA 256

75 rGlnLeuThrGluSerGlyValTyrAlaCysGlyAlaGlyMeta 92
057 GFAGGTCACCAACCAACAACTCAATCAATCGCCGGCGGCG

92 snThrAspArgGly 96
 25/ GCAGCTCACCCAGGACGACACTGGGAGCTACAAGTGTGGCGCTGGGT...A 303

seq_name: qb_est3:AA174968

```
seq_documentation_block:
LOCUS      AAL174968          460 bp      mRNA      EST
DEFINITION ms78h03.rl Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:617717
            5', mRNA sequence.
```

ACCESSION AA174968
 VERSION AA174968.1 GI:1756098
 KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
1 (bases 1 to 460)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
REFERENCE
AUTHORS

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R. The WashU-HHMI Mouse EST Project Unpublished (1996) Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:378541 Seq primer: -28M13 rev2 from Amersham High quality sequence stop: 459. Location/Qualifiers		Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R. The WashU-HHMI Mouse EST Project Unpublished (1996) On Jun 5, 1998 this sequence version replaced gi:3186958. Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:899641 Seq primer: -28m13 rev2 ET from Amersham High quality sequence stop: 141. Location/Qualifiers	
source		source	
1. .460 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:617717" /clone_lib="Soares mouse 3NbMS" /sex="male" /tissue_type="Spleen" /dev_stage="4 weeks" /lab_host="DH10B"		1. .220 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:1366421" /clone_lib="Soares_mammary_gland_NDMMG" /sex="male" /tissue_type="mammary gland" /dev_stage="4 weeks" /lab_host="DH10B"	
note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGGCGCGCGCTGTTTGTGTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."		note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGGCGCGCGCGCTGTTTGTGTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."	
BASE COUNT ORIGIN		BASE COUNT ORIGIN	
108 a 158 c 117 g 77 t		62 a 78 c 42 g 38 t	
alignment_scores: Quality: 157.50 Length: 66 Ratio: 3.150 Gaps: 1 Percent Similarity: 75.758 Percent Identity: 50.000		alignment_scores: Quality: 145.50 Length: 64 Ratio: 3.031 Gaps: 1 Percent Similarity: 75.000 Percent Identity: 48.438	
alignment_block: US-09-135-238B-2_COPY_18_253 x AA174968 .. Align seg 1/1 to: AA174968 from: 1 to: 460		alignment_block: US-09-135-238B-2_COPY_18_253 x AI021484 .. Align seg 1/1 to: AI021484 from: 1 to: 220	
171 ProArgValSerArgAlaSerSerValAlaGlyAspLysProArgThrPh 187 7 CCAGAGTTACAGGACATTTCTGTGTCAGCTACCAAGTCCCGCGCT 56 187 eLeuProSerThrAlaSerLysIleSerAlaLeuGluGlyLeuLeu 204 57 CTTGCCCAACACAGCCTCAAAGACTTCCACTCAGCAAGCA...ATCA 103 204 ysProGlnThrProSerTyrAsnHisHisThrArgLeuHisArgGlnArg 220 104 GGCCCCATAGAGCCGACATATGGGAGAGAGACCCAGCATTATGAGCAAG 153 221 AlaLeuAspTyrGlySerGlnSerGlyArgGluGlyGlnGlyPheHis 236 154 ACAGCCACCATGGCCACACTATGGGAGAGAGACCCAGCGGCTTCAC 201 seq_name: gb_est15:AI021484 seq_documentation_block: LOCUS AI021484 220 bp mRNA EST 16-JUN-1998		173 ValSerArgAlaSerSerValAlaGlyAspLysProArgThrPheLeuP 189 1 GTTACAGAGCATTTCTGTGTCAGCTACCAAGTCCCGCGCTCCTGCC 50 189 oSerThrThrAlaSerLysIleSerAlaLeuGluGlyLeuLysProG 206 51 AGCAACCAACAGCCTCAAGACTTCCACTCAGCAAGCA...ATCAGGCC 97	

206 lnThrProSerTyrAsnHisHisThrArgLeuHisArgGlnArgAlaLeu 222
| | | | | : : : | | | | | | | | | | | | | | | : : : | | | | : : :
98 TAGAGGCCAGCTACAGCCACCACACCCAGACTTCATGAGCAAGGACACGC 147

223 AspTyrGlySerGlnSerGlyArgGluGlyGlnGlyPheHis 236
: : : | | : : | | | | | | | : : | | | : : | | : : | |
148 CACCATGGCCACACTATATGGGAGAGAGAGACCGAGGGGCTTCAC 189

OM of: US-09-135-238B-2_COPY_18_253 to: Issued_Patents_NA.* out_format : pfs
Date: Sep 12, 2000 6:53 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09135238/runat_24082000_091324_17895/app_query.fasta_1.1282
-DB-issued_Patents_NA -QFM=fastap -SUFFIX=p2n.rni
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -CGAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blotum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFM=pfs -NORM=ext -MINLEN=0 -MAXLEN=100000
-USER=US09135238_@CGN1_1_68 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-135-238B-2_COPY_18_253
Query length: 236
Database: Issued_Patents_NA.*
Database sequences: 243080
Database length: 68777915
Search time (sec): 93.130000

score_list:	Sequence	Strd Orig	zScore	EScore	Len	! Documentation
	/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-7		197.00	341.87	1.5e-11	3095
	/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-9		185.50	318.90	2.8e-10	3269
	/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-3		180.50	310.57	8.3e-10	2919
	/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-5		179.50	305.99	1.5e-09	3630
	/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-642-406A-21		166.50	288.79	1.4e-08	1875
	/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-642-406A-20		166.50	281.15	3.6e-08	3517
	/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-1		166.50	281.15	3.6e-08	3517
	/cgn2_6/ptodata/1/ina/5A/COMB.seq:US-08-286-740-3		102.00	148.48	0.8846	6557
	/cgn2_6/ptodata/1/ina/PTCUS.COMB.seq:US-08-286-740-3		102.00	148.48	0.8846	6557
	/cgn2_6/ptodata/1/ina/5C.COMB.seq:US-08-378-939-9		100.50	162.56	0.1454	1617
	/cgn2_6/ptodata/1/ina/5D.COMB.seq:US-08-935-450-10		99.00	166.57	0.0869	914
	/cgn2_6/ptodata/1/ina/5D.COMB.seq:US-08-887-352B-1		98.50	142.52	1.90	6127
	/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-157-101A-9		96.50	156.99	0.2971	1350
	/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-157-101A-6		96.50	155.11	0.3780	1576
	/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-463-587A-24		92.00	142.46	1.92	2178
	/cgn2_6/ptodata/1/ina/5C.COMB.seq:US-08-923-854-24		92.00	142.46	1.92	2178
	/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-923-854-24		92.00	142.46	1.92	2178
	/cgn2_6/ptodata/1/ina/PTCUS.COMB.seq:US-08-923-854-24		92.00	142.46	1.92	2178
	/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-149-099C-3		91.50	123.99	20.45	9209
	/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-476-275-2		91.50	123.99	20.45	9209
	/cgn2_6/ptodata/1/ina/5C.COMB.seq:US-08-476-275-2		91.50	123.99	20.45	9209
	/cgn2_6/ptodata/1/ina/5D.COMB.seq:US-08-819-866-2		91.50	115.21	63.06	18986
	/cgn2_6/ptodata/1/ina/5D.COMB.seq:US-09-023-715-2		91.50	115.21	63.06	18986
	/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-592-126-75		91.00	148.99	0.8283	1083
	/cgn2_6/ptodata/1/ina/6/COMB.seq:US-09-027-449-59		90.50	149.91	0.7363	927
	/cgn2_6/ptodata/1/ina/6/COMB.seq:US-09-027-449-61		90.50	146.16	15.48	6563
	/cgn2_6/ptodata/1/ina/6/COMB.seq:US-09-027-449-68		90.50	123.58	21.57	8120
	/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-836-022A-3		90.50	121.09	29.69	9972
	/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-276-099A-1		90.00	134.51	5.31	3046
	/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-781-890-1		90.00	134.51	5.31	3046
	/cgn2_6/ptodata/1/ina/5D.COMB.seq:US-08-860-882A-28		89.50	150.11	0.7175	2070
	/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-394-326-1		89.00	137.35	3.73	1929
	/cgn2_6/ptodata/1/ina/PTCUS.COMB.seq:US-08-394-326-1		88.50	137.14	3.79	1932
	/cgn2_6/ptodata/1/ina/PTCUS.COMB.seq:US-08-394-326-1		88.50	137.12	3.80	1941
	/cgn2_6/ptodata/1/ina/PTCUS.COMB.seq:US-08-394-326-1		88.50	137.07	3.82	1941
	/cgn2_6/ptodata/1/ina/PTCUS.COMB.seq:US-08-394-326-1		88.50	137.07	3.82	1941
	/cgn2_6/ptodata/1/ina/PTCUS.COMB.seq:US-08-394-326-1		88.50	135.52	4.66	2208
	/cgn2_6/ptodata/1/ina/PTCUS.COMB.seq:US-08-394-326-1		88.50	135.50	4.67	2208
	/cgn2_6/ptodata/1/ina/PTCUS.COMB.seq:US-08-394-326-1		88.50	135.45	4.70	2217
	/cgn2_6/ptodata/1/ina/PTCUS.COMB.seq:US-08-394-326-1		88.50	135.44	4.71	2220

/cgn2_6/ptodata/1/ina/5C.COMB.seq:US-08-344-155C-99 + 88.50 132.56 6.81
/cgn2_6/ptodata/1/ina/5C.COMB.seq:US-08-224-482-3 + 88.50 131.26 8.05
/cgn2_6/ptodata/1/ina/6.COMB.seq:US-09-205-921-1 + 88.50 131.26 8.05
/cgn2_6/ptodata/1/ina/5C.COMB.seq:US-08-551-211-4 + 88.00 138.36 3.24
/cgn2_6/ptodata/1/ina/5C.COMB.seq:US-08-551-211-5 + 88.00 134.57 5.27
seq_name: /cgn2_6/ptodata/1/ina/6.COMB.seq:US-08-434-000A-7
seq_documentation_block:
; Sequence No. 6046037
; Patent No. 08434000A
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application 1
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; TELEX: SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3095 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE: DESCRIPTION: Mouse Polyimmunoglobulin Receptor
; NAME/KEY: Coding Sequence
; LOCATION: 85.....2400
US-08-434-000A-7
alignment_scores:
Quality: 197.00 Length: 148
Ratio: 2.165 Gaps: 5
Percent Similarity: 61.486 Percent Identity: 32.432
alignment_block:
US-09-135-238B-2_COPY_18_253 x US-08-434-000A-7 ..
Align seg 1/1 to: US-08-434-000A-7 from: 1 to: 3095
13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; TELEX: SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; RATIONALE: DESCRIPTION: Rat Polyimmunoglobulin Receptor
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 74...2383
; US-08-434-000A-9

alignment_scores:
    Quality: 185.50      Length: 164
    Ratio: 1.855        Caps: 6
    Percent Similarity: 60.976    Percent Identity: 30.488

alignment_block:
US-09-135-238B-2_copy_18_253 x US-08-434-000A-9 ..
Align seg 1/1 to: US-08-434-000A-9 from: 1 to: 3269

13 GlyGlySerValThrIleLeuValThrLeuProGlu.....Me 25
170 GGTAACTCGGTCTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 219
25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
220 GCACACCCGGAAATACTGTCGCCAAGAGGCC...AACGGCTACTGCG 266
42 IyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 58
267 CAACCTCATCTCTTCAATGGCTACTCTCGAAGGAGTATTCAGGCAGA 316
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
317 GCACGCTCATCACTCACTCCAGAGAATAGCACATTGTGATTAAACATTGC 366
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92
367 ACATCTCACCAGGAGGACACTGGGAGCTACAAGTGTGCTCTGGGT...A 413
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
414 CCACCTACCCAGGAGGCTGTTTTCGATGTCAGCCTGGAGTC..... 454
109 TyrGluProSerTyrGluGlnProMetProGluThrProLysTyrPph 125
455 .....AGCCAGGTTCTGAGTCCCAATGACAC 483
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 142
484 CCATGTC.....TACAAAGG 500
142 heValThrArgValThrProAlaGlnArgGlyLysValProVal 158
501 ACATAGGCGAAGTGTGACCATCGAATGCCGTTTCAAGAGGGGAATGCT 550
159 HisHisSer.SerProThrThrGlnIleThrHisArgPro 171
551 CATAGCAAGAATCCCTGTGTAGGAAGAGAGAGAGAGGCGCT 590

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-3
seq_documentation_block:
; Sequence 9, Application US/08434000A
; Sequence 3, Application US/08434000A
; Patent No. 6046037

181 GCGACTCTGTTTCATPCACGTGCTACTACCCAGACACCTCTGTCAACCG 230
25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
231 GCACACCCGGAAATACTGTCGCCAAGAGGCC...AGCGCATGTGCA 277
42 IyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 58
278 CAACGCTCATCTCTCAATGGCTACCTCTCCAAGGAGTATTCAGGCAGA 327
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
328 GCCAACCTCATCACTCCAGAGAACAACACATTGTGATTAAACATGA 377
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92
378 GCAGCTCACCCAGGAGGACACTGGGAGCTACAAGTGGCCTGGGT...A 424
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
425 CCAGTAACCGAGGCTGTCTTCGATGTGAGCTGGAGTC..... 465
109 TyrGluProSerTyrGluGlnProMetProGluThrProLysTyrPph 125
466 .....AGCCAGGTTCTGAGTGGCGAGTGACAC 494
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 142
495 CCAGTC.....TACAAAGGACATAGCA 520
142 heValThrArgValThrProAlaGlnArgGlyLysValPro 156
521 GAATGTGACCATCAATGCCCTTCAAAAGGAGAATGTGCC 564

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-9
seq_documentation_block:
; Sequence 9, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
```



```
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,406A
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 800
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/591,823
; FILING DATE: 02-OCT-1990
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/427,765
; FILING DATE: 27-OCT-1989
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Logan, April C.
; REGISTRATION NUMBER: 33,950
; REFERENCE/DOCKET NUMBER: 184.2
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
;
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
;
; US-08-642-406A-20

alignment_scores:
  Quality: 166.50      Length: 103
  Ratio: 2.602        Gaps: 4
  Percent Similarity: 62.136  Percent Identity: 42.718

alignment_block:
  US-09-135-238B-2_COPY_18_253 x US-08-642-406A-20 ..
  Align seg 1/1 to: US-08-642-406A-20 from: 1 to: 3517

13 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 25
238 GCGGACTCGGTGTCATCATGTCTACTACCCACACACCTCGTCACCG 287

25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
288 GCACAGCCGGAAGTCTGTGTCGGGAA...GAGGAGAGCGCGCTGCG 334

42 lyThrValValSerThrThrAsnPhelLeLysAlaGluTyrLysGlyArg 58
335 TGACGCTTGCCCTCG...ACCGGCTACAGTCCCGAGGAATCTCCGGGAGA 381

59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
382 GCGAAGCTCACCGACTTCCTCGATAAAGGGAGTTTGGTGACTGTGTA 431

75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
432 CCAACTCACCCAGACGACTCAGGAGGTACAAGTGTGGCTGGGAGTCA 481

92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
482 AC...GGCCGTGGCTGGACTTCGGTGTCAACGTCGTGTCAGCCAGGAG 528

109 TyrGluPro 111

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-1

seq_documentation_block:
; Sequence 1, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application 1
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
;
; SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 124....2445
;
; US-08-434-000A-1

alignment_scores:
  Quality: 166.50      Length: 103
  Ratio: 2.602        Gaps: 4
  Percent Similarity: 62.136  Percent Identity: 42.718

alignment_block:
  US-09-135-238B-2_COPY_18_253 x US-08-434-000A-1 ..
  Align seg 1/1 to: US-08-434-000A-1 from: 1 to: 3517

13 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 25
238 GCGGACTCGGTGTCATCATGTCTACTACCCACACACCTCGTCACCG 287
```


ATTORNEY/AGENT INFORMATION:
 NAME: ERNST, BARBARA G
 REGISTRATION NUMBER: 30,377
 REFERENCE/DOCKET NUMBER: 1808-118
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 783-6040
 TELEFAX: (202) 783-6031
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1617 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FEATURE:
 NAME/KEY: sig_peptide
 LOCATION: 35..92
 FEATURE:
 NAME/KEY: mat_peptide
 LOCATION: 93..1465
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 35..1465
 US-08-378-939-9

alignment_scores:
 Quality: 100.50 Length: 254
 Ratio: 0.817 Gaps: 11
 Percent Similarity: 48.425 Percent Identity: 24.016

alignment_block:

US-09-135-238B-2_COPY_18_253 x US-08-378-939-9 ..

Align seg 1/1 to: US-08-378-939-9 from: 1 to: 1617

```

5 GluValLysValGluGlyLeuGlySerValThrIleLysCysPr 21
|||||
119 GAAGTAAAG.....AAGCCTGGGTCTCGTGACGGTCTCTGCAA 159
21 OleuProglu.....MetHisValArgIleTyrLeuCysArgGluM 35
|||
160 GGCATCTCGAGGACCTTCAGCAACTATGCTATCATCGCTGGGTGGCAGG 209
35 etAlaGlySerGly.....ThrCysGly 42
|||||
210 CCCTGGACAAGGCTGTAGTGGATGGAGGATCATCCCTTTTGGT 259
43 ThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArgVa 59
|||||
260 ACACCAACTACTTCACAGAACTC.....CAGGGCAGAGT 294
59 lThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValThrG 76
|||||
295 CACGATTACCCGGGCAATCCACAGCAGCCACATGGAGTGACTA 344
76 lLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly..... 90
|||||
345 GCCTGAGATCTGAGGACAGCGCGGTGTATTACTGTGCGACAGATCGCTAC 394
91 .....MetAsnThrAspArgGlyLysThr.GlnLysValThrLeuAsnV 105
|||||
395 AGCAGGCAAAATTTGACCGGCCCGGGTGGCTGGTTCGACCCCTGGG 444
105 aHisSerGluTyrGluProSerTrpGluGlnProMetProGluThr 121
|||||
445 CCAGGGACACCTGGTCCACGCTCTCTCAGCCTCCACCAAGGGGCCCATCGG 494
122 ProLysTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSe 138
|||||
495 TCTTCCCTGGCACCTCTCTCCAAAGAGCACCTCTGGGGGCGACAGCGGCC 544

```

```

138 rSerSerLysPheValThrArgValThrThrProAlaGln.....A 152
|||
545 CTGGCTGCCTGG...TCAGGACTACTTCCCGGACGGGTGAGGTGTC 591
152 rgGlyLysValProValHisSerSerProThrThrGlnIleThr 168
|||||
592 GTGGAAGTCTAGGCGCCTGACCAAGCGCGTGCACACCTTCCCGCTGTCC 641
169 HisArgProArgValSerArgAlaSerSerValAlaGlyAspLysProAr 185
|||
642 TACAGTCTCTAGGACTTACTCTCTCAGCAGCGGTGGTGACCGTCCCT.. 689
185 gThrPheLeuProSerThrThrAlaSerLysLysSerAlaLeuGluGlyL 202
|||||
690 .....CCAGCAGCTTGGGACCCAGACCTACATCTGCAACGTGAA 729
202 euLeuLysProGlnThrPro.....SerTyrAsnHis 212
|||
730 TCACAGCCCGACCAACACCAAGGTGGACAAAGATTGACGCCAAATCTT 779
213 HisThrArgLeu...HisArgGlnArgAlaLeuAspTyrGlySerGlnSe 228
|||||
780 GTGACAAACTCACACATGCCCGTCCAGCAGTCCGACCACTGAACTCCTGGGG 829
228 rGlyArgGlu 231
|||
830 GGACCGTCAG 839

```

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-935-450-10

seq_documentation_block:

```

; Sequence 10, Application US/08935450
; Patent No. 5977311
; GENERAL INFORMATION:
; APPLICANT: Nandabalan, Krishnan
; APPLICANT: Yang, MeiJia
; APPLICANT: Schulz, Vincent
; TITLE OF INVENTION: 53BP2 COMPLEXES
; FILE REFERENCE: 7934-054
; CURRENT APPLICATION NUMBER: US/08/935,450
; CURRENT FILING DATE: 1997-09-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 914
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-935-450-10

```

alignment_scores:
 Quality: 99.00 Length: 146
 Ratio: 1.456 Gaps: 9
 Percent Similarity: 46.575 Percent Identity: 29.452

alignment_block:

US-09-135-238B-2_COPY_18_253 x US-08-935-450-10 ..

Align seg 1/1 to: US-08-935-450-10 from: 1 to: 914

```

116 GluProMetProGluThrProLysTrpPheHisLeu..... 127
|||||
232 CAGCCCTGCTCGGAACCCCGGCGCCAGCACCTACATAAGAAACAGCA 281
128 ....ProTyrLeu.....PheGlnMetProAlaTyrAlaSerSerS 140
|||||
282 ACATCCCTGGCTCAAGCGCAATACCAGCACCCCGCCGTCAGCAGCTAC 331
140 erLysPheValThrArgValThr.....ProAlaGlnArg..... 152
|||
332 AGCCCTTCCACAGCGGAGTTACAGCGCCACCTACACACAGGGGAGGT 381

```

```
153 .....GlyLysValProPro.ValHisHisSerSerP 163
    ||| ||||| ::||| |||
382 TACAGCCAGGGTTACAGGCCACCGCTCCACCTCCACCACCACTGC 431
163 roThrThrGln.....IleThrHisArgProArgValSerArg 175
    ||||| ||| ::||| ||||| ::|
432 CTACAATATGGAGTACCGCGGTACACCGCGGCCCTATATACCCAC 481
176 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThr 192
    ::|||
482 CGCCACCCC...CCAGCCGACAGACCTACCTCAGCCCAACTATAACCCAG 528
192 rAlaSerLysIleSerAlaLeuGluGlyLeuLeuLys.....ProG 206
    : |||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
529 TATCAGCAGTATGCCAGCTACGAGCAACAGTACTCAGAACACAGGCCAG 578
206 lnThrProSerThrAspHisHisThrArgLeuHisArgGlnAlaLeu 222
    ::||| |||::: ||::: ||::: ||::: ||::: ||::: ||:::
579 TGGCGCCATACTACGGGAATAC..... 601
223 AspTyrGlySerGlnSerGlyArgGluGlyGlnGly 234
    ||||| ||||| ||||| ||||| |||||
602 GACTACGGGAGCTACTCCGGGAACACACAGGGTGGC 637
```

seq_name: /cgn2_6/ptodata/1/lna/5D_COMB.seq:US-08-887-352B-1

seq_documentation_block:

```
; Sequence 1, Application US/08887352B
; Patent No. 5994511
; GENERAL INFORMATION:
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardiou, John Lowe
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of
; TITLE OF INVENTION: Improving Polypeptides
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/887,352B
; FILING DATE: 03-Jul-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1123
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6127 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Circular
US-08-887-352B-1
```

alignment_scores:

Quality:	98.50	Length:	220
Ratio:	0.947	Gaps:	10
Percent Similarity:	47.273	Percent Identity:	23.636

alignment_block:

US-09-135-238B-2_COPY_18_253 x US-08-887-352B-1 ..

```
Align seg 1/1 to: US-08-887-352B-1 from: 1 to: 6127
13 GlyGlySerValThrIleLysCysProLeuProGluMethHisValArgII 29
    ||||| ||| ::||| |||
1357 GGGGCTCACTCCGTTGCTCCTGTCAGTTTCTGGCTACTCATCACTC 1406
29 e.....TyrLeuCysArgGluMetAlaGlySerGly..... 39
1407 CGGATACAGCTGGAACTGGATCCGTCAGGCCCGGGTAAGGGCCTGGAAT 1456
40 .....ThrCysGlyThrValValSerThrThrAsnPheIleLysAlaGlu 54
    ::| ||| ::||| ||||| ::|
1457 GGGTTGCATCGATTACGTATCAGCGATCGACTAATAT...AACCTAGC 1503
55 TyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLe 71
    ||||| ||| ::||| ||||| ||||| ||||| ||||| |||||
1504 GTCAAGGGCCGTATCACTATAAGTCGCGACGATTCCAAAAACACATCT 1553
71 uValGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysG 88
    ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1554 COTGCAGATGAACAGCTGCGTGTGAGGACACTGCCGTCATATTATGTG 1603
88 lYalaglyMetAsnThrAspArgGlyLys..ThrGlnLysValThrLeuAs 104
    :: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1604 CT.....CGAGGCAGCCACTATTTCGGTCACTGGCAC 1635
104 nValHisSerGluTyrGluProSerTyrGluGluGlnProMetProGluT 121
    :: ||| ||| ||||| ||||| ||||| ||||| |||||
1636 TTCGCGGTGTGGGGTCAAGGAACCCCTGG.....TCACCGTCTCCTCGC 1679
121 hrProLysTrpPheHis..... 127
1680 CTCACCAAGGCCCATCGGTCTCCCTTAGCACCCCTCTCCCAAGACA 1729
128 ProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysPheValTh 144
    ||| ||| ||| ||| ||| |||
1730 CCTCTGGGGGCACAGCGCCCTGGGCTGCCTGGTCA..... 1765
144 rArgValThrThrProAlaGln.....ArgGlyLysValProProV 158
    ||||| ||||| ||| ::| ||||| ||||| |||||
1766 .AGGACTACTTCCCGAACCCGGTGCGTGTGGAACCTCAGCGCCCT 1814
158 alHisHisSerSerProThrThrGlnIleThrHisArgProArgValSer 174
    ::| ::| ::||| ||||| ||||| ||||| ||||| |||||
1815 GACCACGGCGTGACACCTTCCCGGTGCTCTACAGTCCCTCAGGACTCT 1864
175 ArgAlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerTh 191
    |||::: ||| ||| ||||| ||||| ||||| ||||| |||||
1865 ACTCCCTCAGCAGCGGTGGTGACCGTGCCCT.....CCAGCAGC 1902
191 rThrAlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProGlnThrP 208
    ||| ::| ::||| ||||| ||||| ||||| ||||| |||||
1903 TTGGGCACCCAGACCTACATCTGCAACGTGAATCACAGCCCAACACAC 1952
208 roSerTyr 210
    || ::|
1953 CAAGGTGG 1960
```

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-157-101A-9

seq_documentation_block:

```
; Sequence 9, Application US/08157101A
; Patent No. 5808032
; GENERAL INFORMATION:
; APPLICANT: KORIHARA, TATSUYA
; APPLICANT: MATSUKURA, SHIGEKAZU
; APPLICANT: TSURUOKA, NOBUO
; APPLICANT: ARIMA, KENJI
; APPLICANT: NISHIHARA, TATSURO
; TITLE OF INVENTION: ANTI-HBS ANTIBODY GENES AND EXPRESSION
; TITLE OF INVENTION: PLASMIDS THEREFOR
; NUMBER OF SEQUENCES: 9
```


Percent Similarity: 46.816 Percent Identity: 23.221

alignment_block:

US-09-135-238B-2_COPY_18_253 x US-08-463-587A-24 ..

Align seg 1/1 to: US-08-463-587A-24 from: 1 to: 2178

```

13 GlyGlySerValThrIleLeuLysCysProLeuProGluMetHisValArg.. 28
|||||
907 GGGGGCTCACTCCGTTGTCTGTGACGCTTCTGCTTCAACATTAAGA 956
29 .....ileTyLeuCysArgGluMetAlaCysSerGlyThr.....C 41
|||||
957 CACCTATATACACTGGGTGGCTCAGGCCCGCGGTAAAGGCGCTGAATGGG 1006
41 ysGlyThrValValSerThrThrAsnPhelLeLys.....AlaGluTyr 55
|||||
1007 TTCGAAGATTATCCTACGAATGGTTATAGATATGCGGATAGCGTC 1056
56 LysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuVa 72
|||||
1057 AAGGCGCGTTTCACTATAAGCGCAGACACATCCAAACACACAGCCTACCT 1106
72 lGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyA 89
|||||
1107 GCAGATGAACACAGCCTGGCTGTGAGGACACTGCCGTCTATTATTCTTA 1156
89 la.GlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVa 105
|||||
1157 GATGGGAGGGGAGC.....GCTTCTATGCTATGGACTACTGGGGT 1197
105 lHisSerGluTyrGluProSerTyrPgluGluGlnProMetProGluThrp 122
|
1198 CAAGAAACCTGGTCCACCGTCTCCTCGGCCTCCACCAAGGCCCATCGGT 1247
122 roLysTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaser 138
|
1248 CITCCCCCTGGCACCTCTCCAGAGCACCTCTGGGGGCACAGGGGCC 1297
139 SerSerLysPheValThrArgValThrProAlaGln.....Ar 152
|
1298 TGGGTGCTGG...TCAAGGACTACTTCCCGAACCGGTGACGGTGTG 1344
152 gGlyLysValProProValHisHisSerSerProThrThrGlnIleThrH 169
|||||
1345 TGGAACTCAGCGCCCTGACCAGCGCGGTGCACACCTTCCGGCTGTCT 1394
169 isArgProArgValSerArgAlaSerSer.....Val 179
|
1395 ACAGTCTCAGGACTCTACTCCTCAGCAGCGGTGTGACTGTGCCCTCTA 1444
180 AlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLysII 196
|||||
1445 GCAGCTTGGGACCCAGA.....CCTACATCTGCAACGTGAAT 1482
196 eSerAlaLeuGluGlyLeuLeuLysProGlnThrPro..... 208
|
1483 CACAAGC.....CCAGCACACCAAGGTGGACAAGAA 1514
209 .....SerTyrAsnHisHisThrArgLeu.HisArgGlnArgAlaLeu.A 223
|
1515 AGTTAGCCCAATCTTGTGCACAAACTCACAGGGCCCTTCGTTTGTG 1564
223 spTyr..... 224
|
1565 AATATCAGGCCAATCGTCTGACCTGCCTCAACCTCTCTCAATGCTGGC 1614
225 .....GlySerGlnSerGlyArgGluGlyGlnGly 234
|
1615 GCGGCTCTGGTGGTGTCTGGTGGCGGCTCTGAGGGTGGTGGC 1659
```

OM of: US-09-135-238b-2_COPY_18_272 to: GenEmbl:* out_format : pfs
Date: Sep 12, 2000 6:55 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame+p2n.model -DEV=xlp
-Q/cgn2_1/USPIO.spool/US09135238/runat_24082000_091323_17880/app_query.fasta_1.1282
-DB=GenEmbl -OFMT=fastap -SURFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=10.000 -GAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -GAPOP=10.000 -GAPEXT=0.500
-DELOP=6.000 -DELEX=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=ppct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=100000
-USER=US09135238 @CGN1_1_5120 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-135-238b-2_COPY_18_272
Query length: 255
Database: GenEmbl:*
Database sequences: 972840
Database length: 892348106
Search time (sec): 2784.300000

score_list:

Sequence	Strd	Orig	EScore	Len	Documentation
gb_pr2:AF057557	+	1340.00	1822.33	2.1e-93	1339
gb_hg11:AC023534	+	629.50	807.63	6.9e-37	221365
gb_hg22:AC027719	-	252.00	345.25	7.5e-09	58628
gb_to:MMU06431	+	197.00	205.20	4.1e-06	3095
gb_cm:AF091137	+	189.00	245.02	1.5e-05	2848
gb_to:MMP1GR03	+	188.00	260.64	2.0e-06	417
gb_to:MMP1GR03	+	188.00	258.94	2.5e-06	505
gb_to:AR001489	+	188.00	233.02	0.0003	29392
gb_to:RNRG	+	185.50	239.02	0.0003	3269
gb_cm:AB032195	+	181.00	235.39	3.1e-05	2461
gb_pr2:S43433503	+	180.50	232.08	6.1e-06	345
gb_pat:A52094	+	180.50	236.83	4.3e-05	1936
gb_pat:A52093	+	180.50	236.41	4.5e-05	2031
gb_pat:A52091	+	180.50	234.91	5.5e-05	2405
gb_pr2:S62403	+	180.50	233.26	6.8e-05	2898
gb_pr1:HSP1R	+	180.50	233.20	6.8e-05	2919
gb_cm:BP1GR1	+	179.50	229.91	0.0001	3630
gb_cm:BOV1GR	+	179.00	229.48	0.0001	3527
gb_pat:E15470	+	175.50	230.12	0.0001	1911
gb_cm:OC1GECR	+	166.50	212.45	0.0010	3517
gb_hg11:AC023534	-	166.00	175.14	0.1168	221365
gb_to:MMP1GR5	+	146.50	203.22	0.0032	456
gb_pr1:HUM1GRPOLY	+	143.00	193.31	0.0410	2526
gb_to:MMP1GR05	+	137.00	189.78	0.0179	481
gb_pr2:S43433505	+	132.50	186.89	0.0259	333
gb_pr2:S43433504	+	130.00	177.47	0.0867	657
gb_pr2:HS4238323	+	124.00	165.85	0.3848	969
gb_pr2:HS4010101	+	124.00	165.84	0.3853	970
gb_pr2:HS4224864	+	124.00	160.10	0.8046	1857
gb_hg29:AC060797	-	121.00	125.35	69.33	59506
gb_hg15:AC016888	-	121.00	115.80	235.97	175260
gb_to:AF035963	+	118.50	149.83	3.00	2539
gb_pr3:AF0433724	+	117.50	153.48	1.88	1440
gb_cm:RAB1TCBXB	+	114.50	149.95	2.96	1351
gb_pr1:HSNCAMX1	+	111.00	123.15	91.83	16288
gb_pr1:HSNCAME	+	110.50	138.06	13.61	2799
gb_ba1:AP000062	+	110.50	98.27	2.2e+03	251700
gb_hg22:AC027719	+	109.50	109.79	510.16	58628
gb_hg5:AC013305	-	109.50	100.08	1.8e+03	175799

seq_name: gb_pr3:AF057557

seq_documentation_block:

LOCUS AF057557 1339 bp mRNA PRI 04-JUN-1998

DEFINITION Homo sapiens anti-Fas-induced apoptosis (TOSO) mRNA, complete cds.

ACCESSION AF057557

VERSION AF057557.1 GI:3169292

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 1339)

AUTHORS Hitoshi,Y., Lorens,J., Kitada,S.I., Fisher,J., LaBarge,M., Ring,H.Z., Francke,U., Reed,J.C., Kinoshita,S. and Nolan,G.P.

TITLE Toso, a cell surface, specific regulator of Fas-induced apoptosis in T cells

JOURNAL Immunity 8 (4), 461-471 (1998)

MEDLINE 98246046

REFERENCE 2 (bases 1 to 1339)

AUTHORS LaBarge,M. and Hitoshi,Y.

TITLE Direct Submission

JOURNAL Submitted (03-APR-1998) Molecular Pharmacology, Stanford University, 300 Pasteur Dr., Palo Alto, CA 94304, USA

FEATURES

Location/Qualifiers

1..1339

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/map="ig31-q32"

/cell_line="Jurkat"

/cell_type="T cell"

/note="TOSO, 2 cDNA: cloned by PCR using primers designed from the original TOSO.1 cDNA sequence"

1..1339

/gene="TOSO"

20..1192

/gene="TOSO"

/note="transmembrane protein; Ig superfamily member"

/codon_start=1

/product="anti-Fas-induced apoptosis"

/protein_id="AAC18830.1"

/db_xref="GI:3169293"

/translation="MDFWLPLFLPVSGALRILPEVKELGSGVTIKCPLPEMHV RYLCREMAGSGTCGTVVSTNFIKAEYKGVTLKQYPRKNLFLVEVLTQTESGVY ACAGMNTDRGKTQKVLNVHSEYEPSEWQPMETPKWHLPLFQMPAYASSKFFV TRVTPAQRGVPPVHSSPTQIHRPRVSRSSVAGDKPRTFLPTTAKSLALEG LLKPTQSYNHTLRHQALDYQSREGQGFHLIPTLGLLALGLGVVKKRAV ERRKLSRRRLAVRRALESQRPRSQNNIYSACPRRAGDAAGTGEAP VPGCAPLPPALQVSESPWLHAPSLKTSCEVSLYHOPAAAMEDSDDDYINVPA"

127

/gene="TOSO"

/note="occurs in TOSO.1 cDNA"

/replace="g"

309 a 432 c 340 g 258 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 1340.00

Length: 255

Ratio: 5.255

Gaps: 0

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

US-09-135-238B-2_COPY_18_272 x AF057557 ..

Align seg 1/1 to: AF057557 from: 1 to: 1339

```
1  ArgIleLeuProGluValLysValGluGluLeuGlyGlySerValTh 17
|||||
71  AGGATCCTCCAGAAAGTAAAGGTAGAGGGGAGCTGGCGGATCAGTTAC 120
|||||
17  rIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 34
|||||
121 CATCAAGTCCCACTCTCTGGAATGCATGTGAGGATATATCTGTCGGG 170
|||||
34  luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
|||||
171 AGATGGCTGGATCTGGAACATGTGTACCGTGTATCCACCACTTC 220
|||||
51  IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67
|||||
221 ATCAGGCGAGATACAGGCGCGAGTTACTCTGAGCATATCCACGCAA 270
|||||
67  sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84
|||||
271 GAATCTGTCTCTAGTGGAGTAAACACAGCTGACAGAAAGTGACAGCGAG 320
|||||
84  alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
|||||
321 TCTATGCTCGGAGCGGCATGACACAGACCGGGGAAGACCCAGAA 370
|||||
101 ValThrLeuAsnValHisSerGluTyrGluProSerTrpGluGluGlnPr 117
|||||
371 GTACCCCTGAATGTCCACAGTGAATACGAGCATCATCTGGGAAGACGAC 420
|||||
117 oMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMetP 134
|||||
421 AATGCTCGAGACTCCAAATGGTTTCATCTGCCCTATTGTTCAGATGC 470
|||||
134 roAlaTyrAlaSerSerSerLysPheValThrArgValThrThrProAla 150
|||||
471 CTGCATATGCCAGTCTCTCCAAATTCGTAACCAAGATTACACACAGCT 520
|||||
151 GlnArgGlyLysValProProValHisHisSerSerProThrThrGlnIl 167
|||||
521 CAAGGGGCGAGGTCCTCCAGTTCACACACTCTCCCCACCCACCAAT 570
|||||
167 eThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLysP 184
|||||
571 CACCCACCGCCCTCGAGTGTCCAGAGCATCTTCAGTAGGAGGTGACAGC 620
|||||
184 roArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuGlu 200
|||||
621 CCGGAACCTTCTGCCATCCACTACAGCTCATAAATCTCAGCTCTGGAG 670
|||||
201 GlyLeuLeuLysProGlnThrProSerTyrAsnHisHisThrArgLeuHi 217
|||||
671 GGCTGCTCTCAAGCCCCAGAGCCCGAGCTACAAACCCACACAGGCTGCA 720
|||||
217 sArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGlyGlnG 234
|||||
721 CAGGCAGAGACACTGGACTATGGCTACAGCTCTGGGAGGAAGGCCAAG 770
|||||
234 lyPheHisIleLeuIleProThrIleLeuGlyLeuPheLeuAlaLeu 250
|||||
771 GATTTACATCTGATCCGAGACCATCTCTGGGCCCTTTTCTCTGTGCACTT 820
|||||
251 LeuGlyLeuValVal 255
|||||
821 CTGGGCTGGTGGTG 835
|||||
```

seq_name: gb_htg11:AC023534

seq_documentation_block:

LOCUS AC023534 221365 bp DNA HTG 02-MAR-2000

DEFINITION

Homo sapiens clone RP11-462N18, *** SEQUENCING IN PROGRESS ***, 61

unordered pieces.

ACCESSION AC023534

VERSION 2 GI:7143452

KEYWORDS HTG: HTGS_PHASE1.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 221365)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome, clone RP11-462N18

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 221365)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bada,F., Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M., Feneator,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Larcocque,K., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE

Direct Submission

JOURNAL Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Mar 2, 2000 this sequence version replaced gi:6978210.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5244

Center clone name: 462_N_18

* NOTE: This is a 'working draft' sequence. It currently

* consists of 61 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1108: contig of 1108 bp in length

* 1109 1208: gap of 100 bp

* 1209 2232: contig of 1084 bp in length

* 2293 2392: gap of 100 bp

* 2393 3673: contig of 1281 bp in length

* 3674 3773: gap of 100 bp

* 3774 4947: contig of 1174 bp in length

* 4948 5047: gap of 100 bp

* 5048 6104: contig of 1057 bp in length

* 6105 6204: gap of 100 bp

* 6205 7259: contig of 1055 bp in length

* 7260 7359: gap of 100 bp

* 7360 8667: contig of 1308 bp in length

* 8668 8767: gap of 100 bp

```
*      8768      10566: contig of 1799 bp in length
*      10567      10666: gap of 100 bp in length
*      10867      12173: contig of 1507 bp in length
*      12174      12273: gap of 100 bp
*      12274      13643: contig of 1370 bp in length
*      13644      13743: gap of 100 bp
*      13744      15215: contig of 1472 bp in length
*      15216      15315: gap of 100 bp
*      15316      16609: contig of 1294 bp in length
*      16610      16709: gap of 100 bp
*      16710      18085: contig of 1376 bp in length
*      18086      18185: gap of 100 bp
*      18186      19668: contig of 1483 bp in length
*      19669      19768: gap of 100 bp
*      19769      22083: contig of 2315 bp in length
*      22084      22183: gap of 100 bp
*      22184      24052: contig of 1869 bp in length
*      24053      24152: gap of 100 bp
*      24153      25902: contig of 1750 bp in length
*      25903      26002: gap of 100 bp
*      26003      27639: contig of 1637 bp in length
*      27640      27739: gap of 100 bp
*      27740      29377: contig of 1638 bp in length
*      29378      29477: gap of 100 bp
*      29478      30910: contig of 1433 bp in length
*      30911      31010: gap of 100 bp
*      31011      31225: contig of 215 bp in length
*      31226      31325: gap of 100 bp
*      31326      32890: contig of 1565 bp in length
*      32891      32990: gap of 100 bp
*      32991      34812: contig of 1822 bp in length
*      34813      34912: gap of 100 bp
*      34913      36817: contig of 1905 bp in length
*      36818      36917: gap of 100 bp
*      36918      38575: contig of 1658 bp in length
*      38576      38675: gap of 100 bp
*      38676      41309: contig of 2634 bp in length
*      41310      41409: gap of 100 bp
*      41410      43344: contig of 1935 bp in length
*      43345      43444: gap of 100 bp
*      43445      45853: contig of 2409 bp in length
*      45854      45953: gap of 100 bp
*      45954      48434: contig of 2481 bp in length
*      48435      48534: gap of 100 bp
*      48535      51026: contig of 2492 bp in length
*      51027      51126: gap of 100 bp
*      51127      54369: contig of 3243 bp in length
*      54370      54469: gap of 100 bp
*      54470      56661: contig of 2192 bp in length
*      56662      56761: gap of 100 bp
*      56762      59097: contig of 2336 bp in length
*      59098      59197: gap of 100 bp
*      59198      61551: contig of 2354 bp in length
*      61552      61651: gap of 100 bp
*      61652      64257: contig of 2606 bp in length
*      64258      64357: gap of 100 bp
*      64358      67511: contig of 3154 bp in length
*      67512      67611: gap of 100 bp
*      67612      70329: contig of 2718 bp in length
*      70330      70429: gap of 100 bp
*      70430      73572: contig of 3143 bp in length
*      73573      73672: gap of 100 bp
*      73673      77203: contig of 3531 bp in length
*      77204      77303: gap of 100 bp
*      77304      80599: contig of 3296 bp in length
*      80600      80699: gap of 100 bp
*      80700      84046: contig of 3347 bp in length
*      84047      84146: gap of 100 bp
*      84147      88357: contig of 4211 bp in length
*      88358      88457: gap of 100 bp
*      88458      92454: contig of 3997 bp in length
*      92455      92554: gap of 100 bp
*      92555      96892: contig of 4338 bp in length
```

```
*      96893      96992: gap of 100 bp
*      96993      101334: contig of 4342 bp in length
*      101335      101434: gap of 100 bp
*      101435      105438: contig of 4004 bp in length
*      105439      105538: gap of 100 bp
*      105539      110413: contig of 4875 bp in length
*      110414      110513: gap of 100 bp
*      110514      115994: contig of 5481 bp in length
*      115995      116094: gap of 100 bp
*      116095      120447: contig of 4353 bp in length
*      120448      120547: gap of 100 bp
*      120548      124483: contig of 3936 bp in length
*      124484      124583: gap of 100 bp
*      124584      130249: contig of 5666 bp in length
*      130250      130349: gap of 100 bp
*      130350      135462: contig of 5113 bp in length
*      135463      135562: gap of 100 bp
*      135563      139847: contig of 4285 bp in length
*      139848      139947: gap of 100 bp
*      139948      146731: contig of 6784 bp in length
*      146732      146831: gap of 100 bp
*      146832      151469: contig of 4638 bp in length
*      151470      151569: gap of 100 bp
*      151570      159901: contig of 8332 bp in length
*      159902      160001: gap of 100 bp
*      160002      168308: contig of 8307 bp in length
*      168309      168408: gap of 100 bp
*      168409      176195: contig of 7787 bp in length
*      176196      176295: gap of 100 bp
*      176296      189303: contig of 13008 bp in length
*      189304      189403: gap of 100 bp
*      189404      202135: contig of 12732 bp in length
*      202136      202235: gap of 100 bp
*      202236      221365: contig of 19130 bp in length.

FEATURES             Location/Qualifiers
     source            1..221365
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone_lib="RPC1-11 Human Male BAC"
     misc_feature       1..1108
                        /note="assembly_fragment"
     misc_feature      1209..2292
                        /note="assembly_fragment"
     misc_feature      2393..3673
                        /note="assembly_fragment"
     misc_feature      3774..4947
                        /note="assembly_fragment"
     misc_feature      5048..6104
                        /note="assembly_fragment"
     misc_feature      6205..7259
                        /note="assembly_fragment"
     misc_feature      7360..8667
                        /note="assembly_fragment"

alignment_scores:
    Quality: 629.50      Length: 476
    Ratio: 3.459        Gaps: 6
    Percent Similarity: 38.235      Percent Identity: 34.244

alignment_block:
US-09-135-238B-2_COPY_18_272 x AC023534 ..
Align seg 1/1 to: AC023534 from: 1 to: 221365
1 ArgileLeuProGluValLysValGluGlyGluLeuGlyGlySerValTh 17
|||||
33190 AGGATCCTCCACAGGTAAAGGTAGAGGGGAGCTGGCGGATCAGTTAC 33239
17 rlleLysCysProLeuProGluMethHisValArgIleTyrLeuCysArg 34
|||||
33240 CATCAAGTGCCCACTTCCTGAAATGCATGTGAGGATATATCTGTGCCGG 33289
34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
```

|||||
33290 AGATGGCTGGATCTGGAAATGTGTACCGTGGTATCCACCAACTTC 33339
51 IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67
|||||
33340 ATCAGGCGAGATACAGAGCGCGAGTTACTCTGAGCAATACCCAGCAA 33389
67 sAsnLeuPheLeuValGlnValThrGlnLeuThrGluSerAspSerGlyV 84
|||||
33390 GAATCTGTCTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGGAG 33439
84 alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
|||||
33440 TCTATGCTCGCGAGCGGCATGAACACAGACCGGGGAAGACCCAGAA 33489
101 ValThrLeuAsnValHis 106
|||||
33490 GTCACCCCTGAATGTGCACAGCTGGTAGGTTCTCCGCTGATTGGAGGCTCA 33539
106 106
33540 GACTACCCAGAAATGTTTCACTTGGGAGATAGGAACAGACATCTACT 33589
106 106
33590 GTGACAGAGTCTAGACTCTTACACTGTGGCCATATAAGCCACATAATGA 33639
106 106
33640 TTGAGATCCATGAACACCTCATAGTCAATTACTCTCATGTCGCACACAA 33689
106 106
33690 TTTTCTACCAACATGTATGAAAAAGAACAGATAGGTGCTCTCTTTGTAT 33739
106 106
33740 TGTTCATCTACCCAGGCATGAGGGAGTAGAGGTTGAGATTGAGGAG 33789
106 106
33790 AGGCTCTGAGATGGAGGAATTGGAAGCAAGAAATAACTATCCAAAG 33839
106 106
33840 TAATTGTCCTCATGAGAAATAAGGAAGACACTAATTTATCCAGCCACAT 33889
106 106
33890 TTATAGCCCAATCTTATTACTCTGTGACCTGATCCCTTTATTTCTGGAA 33939
106 106
33940 CCCCCTTTCTCTTCTTCTCCACTGAGTTCCTCCACACATATGACAGCT 33989
106 106
33990 AATTCAACAGCCTAGATACCAATCCCAATGACCTCTGCTGTTTTTGAGGAATG 34039
106 106
34040 TTGACTGGACTAGGGGACCACTATTTCTCCAGGGAGGCCCAACCCA 34089
106 106
34090 GCTCCTTCTCTGCTTTCTTTCTAGTGAGCCCTGCCCTCATCCCTCACTGTC 34139
106 106
34140 ATCCACTTTGCTGAATAGCTTCACTTACAGTTTACACCAAGCCTCAGTCGCC 34189
107 107
107Ser..GluTyrGluProS 112 *
::: |||||||

34190 CCTAGAGACCACCAATATTCTCTCTTCCATAGAATACGAGCCAT 34239
112 erTTPGluGluGlnProMetProGluThrProLysTyrPheHisLeuPro 128
|||||
34240 CATGGGAGAGCAGCCATGCTGAGACTCCAATGGTTTCATCTGCCC 34289
129 TyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysPheValThrAr 145
|||||
34290 TATTGTTCAGATGCTTCATATGCGAGTTCTTCCAAATTCGTAACCCAG 34339
145 g...ValThrThrProAlaGlnArgGlyLysValPro..... 156
|||||
34340 AGGTCAGTTCACCATAGCAGGGGTAGGAGGTGGGTGAGATGCTGTGCC 34389
157ProValHisHisSerSerPro 163
|||||
34390 AATGTAAATGGAGGGGGTCTATGTTTCAGTTTCAT..... 34427
164 ThrThrGlnIleThrHisArgProArgValSerArgAlaSerSerValAl 180
|||||
34428AGATGGCTGGCAGAGCTGAGCTCTTAGTGTGTATTCCTC 34471
180 aGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLysIles 197
::: |||
34472 TGCCTTTAGAGAGGCGACATACCTCCCTGAGATCTGCGGTCCATATCCTT 34521
197 erAlaLeuGluGlyLeuLeuLysProGlnThrProSerTyrAsnHisHis 213
|||||
34522 CAAGCCTGTCGACAGCTCAGTCCCTGCTCTAGCTGCAGAACTTACAT 34571
214ThrArgLeuHisArgGlnArg 220
|||||
34572 ATTACCACGAGGTGGCACCACCAAGCAA 34598
seq_name: gb_htg22.AC027719
seq_documentation_block:
LOCUS AC027719 58628 bp DNA HTG 01-APR-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-258P9 map 8, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC027719
VERSION AC027719.1 GI:7382531
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 58628)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-258P9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 58628)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Headford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienda,V., Morrow,J.,
Murphy,T., Naylor,J., O'Connor,C.H., O'Connell,P.,
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7449
Center clone name: 258_P_9

* NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 671: contig of 671 bp in length
672 771: gap of 100 bp
772 1469: contig of 698 bp in length
1470 1569: gap of 100 bp
1570 2261: contig of 692 bp in length
2262 2361: gap of 100 bp
2362 3051: contig of 690 bp in length
3052 3151: gap of 100 bp
3152 3816: contig of 665 bp in length
3817 3916: gap of 100 bp
3917 4597: contig of 681 bp in length
4598 4697: gap of 100 bp
4698 5395: contig of 698 bp in length
5396 5495: gap of 100 bp
5496 6185: contig of 690 bp in length
6186 6285: gap of 100 bp
6286 6969: contig of 684 bp in length
6970 7069: gap of 100 bp
7070 7780: contig of 711 bp in length
7781 8590: contig of 710 bp in length
8591 8690: gap of 100 bp
8691 9372: contig of 682 bp in length
9373 9472: gap of 100 bp
9473 10162: contig of 690 bp in length
10163 10262: gap of 100 bp
10263 10944: contig of 682 bp in length
10945 11044: gap of 100 bp
11045 11747: contig of 703 bp in length
11748 11847: gap of 100 bp
11848 12547: contig of 700 bp in length
12548 12647: gap of 100 bp
12648 13336: contig of 689 bp in length
13337 13436: gap of 100 bp
13437 14103: contig of 667 bp in length
14104 14203: gap of 100 bp
14204 14872: contig of 669 bp in length
14873 14972: gap of 100 bp
14973 15682: contig of 710 bp in length
15683 15782: gap of 100 bp
15783 16491: contig of 709 bp in length
16492 16591: gap of 100 bp
16592 17290: contig of 699 bp in length
17291 17390: gap of 100 bp
17391 18088: contig of 698 bp in length
18089 18188: gap of 100 bp
18189 18883: contig of 695 bp in length

* 18884 18983: gap of 100 bp
18884 19671: contig of 688 bp in length
19672 19771: gap of 100 bp
19772 20413: contig of 642 bp in length
20414 20513: gap of 100 bp
20514 21196: contig of 683 bp in length
21197 21296: gap of 100 bp
21297 21992: contig of 696 bp in length
21993 22092: gap of 100 bp
22093 22804: contig of 712 bp in length
22805 22904: gap of 100 bp
22905 23604: contig of 700 bp in length
23605 23704: gap of 100 bp
23705 24402: contig of 698 bp in length
24403 24502: gap of 100 bp
24503 25191: contig of 689 bp in length
25192 25291: gap of 100 bp
25292 25974: contig of 683 bp in length
25975 26074: gap of 100 bp
26075 26769: contig of 695 bp in length
26770 26869: gap of 100 bp
26870 27572: contig of 703 bp in length
27573 27672: gap of 100 bp
27673 28373: contig of 701 bp in length
28374 28473: gap of 100 bp
28474 29165: contig of 692 bp in length
29166 29265: gap of 100 bp
29266 29956: contig of 691 bp in length
29957 30056: gap of 100 bp
30057 30741: contig of 685 bp in length
30742 30841: gap of 100 bp
30842 31519: contig of 678 bp in length
31520 31619: gap of 100 bp
31620 32323: contig of 704 bp in length
32324 32423: gap of 100 bp
32424 33110: contig of 687 bp in length
33111 33210: gap of 100 bp
33211 33902: contig of 692 bp in length
33903 34002: gap of 100 bp
34003 34708: contig of 706 bp in length
34709 34808: gap of 100 bp
34809 35500: contig of 692 bp in length
35501 35600: gap of 100 bp
35601 36300: contig of 700 bp in length
36301 36400: gap of 100 bp
36401 37097: contig of 697 bp in length
37098 37197: gap of 100 bp
37198 37900: contig of 703 bp in length
37901 38000: gap of 100 bp
38001 38695: contig of 695 bp in length
38696 38795: gap of 100 bp
38796 39487: contig of 692 bp in length
39488 39587: gap of 100 bp
39588 40291: contig of 704 bp in length
40292 40391: gap of 100 bp
40392 41101: contig of 710 bp in length
41102 41201: gap of 100 bp
41202 41886: contig of 685 bp in length
41887 41986: gap of 100 bp
41987 42686: contig of 700 bp in length
42687 42786: gap of 100 bp
42787 43484: contig of 698 bp in length
43485 43584: gap of 100 bp
43585 44276: contig of 692 bp in length
44277 44376: gap of 100 bp
44377 45078: contig of 702 bp in length
45079 45178: gap of 100 bp
45179 45869: contig of 691 bp in length
45870 45969: gap of 100 bp
45970 46664: contig of 695 bp in length
46665 46764: gap of 100 bp
46765 47456: contig of 692 bp in length
47457 47556: gap of 100 bp

* 47557 48249: contig of 693 bp in length
* 48250 48349: gap of 100 bp
* 48350 49042: contig of 693 bp in length
* 49043 49142: gap of 100 bp
* 49143 49843: contig of 701 bp in length
* 49844 49943: gap of 100 bp
* 49944 50640: contig of 697 bp in length
* 50641 50740: gap of 100 bp
* 50741 51438: contig of 698 bp in length
* 51439 51538: gap of 100 bp
* 51539 52231: contig of 693 bp in length
* 52232 52331: gap of 100 bp
* 52332 53020: contig of 689 bp in length
* 53021 53120: gap of 100 bp
* 53121 53857: contig of 737 bp in length
* 53858 53957: gap of 100 bp
* 53958 54652: contig of 695 bp in length
* 54653 54752: gap of 100 bp

alignment_scores:
Quality: 252.00 Length: 48
Ratio: 5.250 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-135-238B-2_COPY_18_272 x AC027719/rev ..

Align seg 1/1 to reverse of: AC027719 from: 1 to: 58628

1 Arg1leLeuProGluValLysValGluGlyLeuLeuGlySerValTh 17
40580 AGGATCTCCCAAGAGTAAGAGTAGAGGGGAGCTGGCGGATCATTTAC 40531

17 r1leLysCysProLeuProGluMetHisValArg1leTyLeuCysArg 34
40530 CATCAAGTCCCACTCTCTGAAATGCATGTGAGGATATCTGTGCGGG 40481

34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThr 48
40480 AGATGGCTGGATCTGGACATGTGTACCGTGTATCCACCACC 40437

seq_name: gb_ro:MMU06431

seq_documentation_block: 3095 bp mRNA ROD 26-MAY-1995
LOCUS MMU06431
DEFINITION Mus musculus B6/CBAF1J polymeric immunoglobulin receptor mRNA,
complete cds.

ACCESSION U06431

VERSION U06431.1 GI:458421

KEYWORDS mouse.

SOURCE Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE Piskurich,J.F., Blanchard,M.H., Youngman,K.R., France,J.A. and

Kaetzel,C.S.

TITLE Molecular cloning of the mouse polymeric Ig receptor. Functional

regions of the molecule are conserved among five mammalian species

J. Immunol. 154 (4), 1735-1747 (1995)

J. Immunol. 154 (4), 1735-1747 (1995)

95138517

2 (sites)

REFERENCE Piskurich,J.F., Hsieh,M., Youngman,K.R. and Kaetzel,C.S.

AUTHORS Molecular cloning of mouse polymeric Ig receptor

J. Immunol. 150, 38A-38A (1993)

J. Immunol. 150, 38A-38A (1993)

REFERENCE 3 (bases 1 to 3095)

AUTHORS Piskurich,J.F.

TITLE Molecular Cloning and Regulation of the Polymeric Immunoglobulin

Receptor

Thesis (1994) Pathology, Case Western Reserve University

4 (bases 1 to 3095)

REFERENCE Piskurich,J.F.

AUTHORS Direct Submission

JOURNAL Submitted (08-FEB-1994) Janet F. Piskurich, Case Western Reserve
University, Pathology, Biomedical Research Building, Cleveland, OH
44060, USA

FEATURES

source

1..3095 Location/Qualifiers
/organism="Mus musculus"
/strain="B6/CBAF1J (C57 Black 6 X CBA)"
/db_xref="taxon:10090"
/clone="G20"
/clone_lib="catalog #935302, Stratagene Cloning Systems,
La Jolla, CA"
/sex="female"
/tissue_type="liver"
..84
5'UTR 85..138
CDS 85..2400
/codon_start=1
/function="binds and transports polymeric immunoglobulin"
/product="polymeric immunoglobulin receptor"
/protein_id="AAA67440.1"
/db_xref="GI:458422"
/translation="MRLVLTLLVTFVSGVSTKSPIFGPQEVSSIIEGDSVITCYVPD
TSVNRHTRKYKCGAGSCMCTLISSNGYLSKEYSGRANLINFENNFTVINEOLTQ
DDTGYKCGLGTSNGLSDFSVLSVQYPELSDTHVTYTKDIGNVTIECPKRENVP
SKSLCKRTNQSCELVIDSTEKVPSTIGRAKLFMKGIDTLTFYVNIHSLTHNDAGLY
ICQAGEGFSADKKNDLQVLAPELLELYKDLRSVTFECDDGREGVANEAKILCRANKE
TCQVITLTKRDPDFEGRILITPKDDNGRFSVLTGLRKEDGAGHCCGSGGLPQE
GWPQTQWLFNEESTIPNRRSVKVTGGSVAIACPNPKSSSLKYNCRWEGDNG
HCPALVGTQAOVEYEGRLALFDQPGNGTYVILNQLTDEDAGFYMCLTNGDSWRIT
TLEQVAETREPNTVPONATVLTGETFTYSCHVPCKFYSOEYKWKSNKSGCHIL
PSHDEGARQSSVSCDSSQSLMTLNPVSKDEGHWGCVKGOTYGETTATYIAVEE
RTGSSHWNTIDANAKVALEEVVDSISEKENKAIPNPGFANERETQWVRQAO
ENRAGDAGSADGSGSSSKVLFSTLPLGLVLAVGAIAVAVARHRRKNDVRMSIS
SYRTDISMADFKNSRDGNDNMFGASPDQTVIEKQDEIVTITTECTAEPEESKAKAR
SSKEEADWYSAFLQSSITAAQVHDGPOEA"

sig_peptide

misc_feature

misc_feature

misc_feature

3'UTR

BASE COUNT 861 a 796 c 784 g 654 t

ORIGIN

alignment_scores:

Quality: 197.00 Length: 148

Ratio: 2.165 Gaps: 5

Percent Similarity: 61.486 Percent Identity: 32.432

alignment_block:

US-09-135-238B-2_COPY_18_272 x MMU06431 ..

Align seg 1/1 to: MMU06431 from: 1 to: 3095

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25

181 GGGGACTCTGTTCCATCAGTGTACTACCCAGACACCTGTCTCAACCG 230

25 tHisValArgIleTyLeuCysArgGluMetAlaGlySerGlyThrCysG 42

231 GCACACCGGAAATACTGGTGGCGCAAGAGGCC...AGCGGATGTGCA 277

42 lyThrValValSerThrThrAsnPhelIleLysAlaGluTyIleLysGlyArg 58

278 CAACGCTCATCTCTCAATGGTACCTCTCAAGAGGAGTATTCAGGCAGA 327

59 ValThrLeuLysGlnTyProArgLysAsnLeuPheLeuValGluValTh 75

328 GCCAACCTCATCACTCCCGAGAGAACACACATTTGTGATTAACATTGA 377


```

75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
378 GCAGCTCACCGACGACACTGGAGCTACAGTGTGCCCTGGGT...A 424
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
425 CCAGTAACCGAGCGCTCTCTCGATGTCAGCGTGGAGGTC..... 465
109 TyrGluProSerTrpGluGlnProMetProGluThrProLysTrpPh 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
466 .....ACCGAGGTCCTGAGTTCCGAGTGACAC 494
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 142
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
495 CCAGCTC.....TACACAAAGGACATAGGCA 520
142 heValThrArgValThrThrProAlaGlnArgGlyLysValPro 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
521 GAAATGTGACCATTAATGCCCTTTCAAAGGGAGAATGTTCC 564

```

seq_name: gb_om:AF091137

seq_documentation_block: 2848 bp mRNA MAM 01-JUL-1999
LOCUS AF091137 Trichosurus vulpecula polymeric-immunoglobulin receptor precursor
DEFINITION Trichosurus vulpecula (PIGR) mRNA, complete cds.

ACCESSION AF091137

VERSION AF091137.1 GI:5305512

KEYWORDS Trichosurus vulpecula.

SOURCE Trichosurus vulpecula.

ORGANISM Trichosurus vulpecula.

REFERENCE 1 (bases 1 to 2848)

AUTHORS Adamski, F.M. and Demmer, J.

TITLE Cloning and characterization of pIGR and J chain of the marsupial,

Trichosurus vulpecula (brush-tailed possum)

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2848)

AUTHORS Adamski, F.M. and Demmer, J.

TITLE Direct Submission

JOURNAL Submitted (09-SEP-1998) Dairy Science, AgResearch Ruakura, East

Street, Hamilton, New Zealand

Location/Qualifiers

1. .2848

/organism="Trichosurus vulpecula"

/db_xref="taxon:9337"

/tissue_type="mammary gland"

1. .2848

/gene="PIGR"

149. .2350

/gene="PIGR"

/codon_start=1

/product="polymeric-immunoglobulin receptor precursor"

/protein_id="AAB41688.1"

/db_xref="GI:5305513"

/translation="MAFFLACLLALLPVVSMKSPIFGKQVTVGGSVSIQCFVPS

SVNRGRKVFCLNLRQSCFTIVSSNGFVSERFSGRKLTNFPGNNSFLIQISLEKE

DGLYKGLGTSNRGLSEDTILEVSGPNLPNNTVEIVTEVKTVSGINCPQEQNTQD

RKFLCKDGSICALVIDSQGVGPDITGRARLSISGTSRVRVFTISQIKRQDVGMVY

CGVEDSDTGIQKNVDLLEPELIIYALGSGVTLNLCALGSLVASVPKFLCKMRAK

ETCDLVNKSFTNNATHGRILFSTHTETGFSIMITQVRKEDGVYHCGAEDGQPS

EEGPTRALQIFVEETVPSPLVWVPGSGSVTITCHYDPKNNNTLYKWKWEGSSH

CTKLVDLGVWVDSYEGRVALWDEPENGIFTVLNLTQPDAGYVWCLNSGHNKRS

VKLEINDGQPLLLAPKYTVLAQLQSILTISCHFPCKFYSEKYKCKWNSGCGETLPTQ

EGSQAFVDCNQNSRVSLFNSVRDHGHWGVCVKNGQNYGETIAVSVEEVEVG

NALOPTNDAVEPKVKGKEIYPTDLGSTEHSRSGSVLFLVLALVLTVA

VALGITKARWRFSRVSGSYRTDLSMSELENPNRQFGANMDASVQETTLGDEE

IITQTEPESIQPKQVKSKEEDAMVTAFLQSNMNAEIPLDPSNA"

BASE COUNT 769 a 690 c 722 g 667 t

ORIGIN

alignment_scores:
Quality: 189.00 Length: 187
Ratio: 1.734 Gaps: 7
Percent Similarity: 58.289 Percent Identity: 31.016

alignment_block:

US-09-135-238B-2_COPY_18_272 x AF091137 ..

Align seg 1/1 to: AF091137 from: 1 to: 2848

```

7 LysValGluGlyGluLeuLeuGlyGlySerValThrIleLysCysProLeuPr 23
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
224 CAAGTGACAGAGTGGGAAGGAGGATCTGTCCATCCAGTGCTTCTACCC 273
|
23 O.....GluMetHisValArgIleTyrLeuCysArgGluMetA 36
|
274 ATCTCTTTCAGTCAATCGCATGCCGCCCAATACTTTTGTCTTCAGAAATC 323
|
36 laGlySerGlyThrCysGlyThrValValSerThrThrAsnPheIleLys 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 TAAGGCAA...TCCTGTGAACCATAGTCTCTCAATGTTTGTCTCT 370
|
53 AlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLe 69
:|||||:|||||:|||||:|||||:|||||:|||||:
371 GAGAGATTAGTGGGAGAGCCAAACATCCCAATTCCCGAGGGAATACTC 420
|
69 uPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyrA 86
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
421 CTTTTTAATACAGATTTTCAGAGCTGGAGAAGAACATTTGGGCTCTACA 470
|
86 laCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThr 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
471 AGTGTGCTCTGGG...ACGAGTAATCGGGTCTGTCTTTGACATCACC 517
|
103 LeuAsnValHisSerGluTyrGluProSerTrpGluGluGlnProMetPr 119
|||||:|||||:|||||:|||||:|||||:|||||:
518 TTGGAGGTG.....GGAGAGGTCCTCCATCTCCC 546
|
119 oGlu..ThrProLysTrpPheHisLeuProTyrLeuPheGlnMetProAla 135
:|||||:|||||:|||||:|||||:|||||:|||||:
547 TACAACACTGAGGTCTACCTGCTCACTGAAGTGGGTAAACAGTCTCCATCA 596
|
136 TyrAlaSerSerLysPheValThrArgValThrThrProAlaGlnAr 152
|||||:|||||:|||||:|||||:|||||:|||||:
597 ACTGCCCTTTCCAGGACACAAATACACAGGACAGGAAATCTTATCAAG 646
|
152 gGlyLysValProProValHisHisSerSer..... 162
:|||||:|||||:|||||:
647 AAGGATGGCGAAAGCTGTGCACCTGGTCATCGATTCTCAGGAACAAGTGGG 696
|
163 ..ProThrThrGlnIleThrHisArgProArgValSerArgAlaSerSer 178
|||||:|||||:|||||:|||||:|||||:
697 CCTGAGCTACACAG.....GAAGAGCCCGCCTCAGCATCAGCGGCACCA 740
|
179 ValAlaGly 181
:|||||:
741 GCAGTAGGG 749

```

seq_name: gb_ro:MMP1GR03

seq_documentation_block:

LOCUS MMP1GR03 417 bp DNA ROD 16-APR-1999

DEFINITION Mus musculus polymeric immunoglobulin receptor (pIGR) gene, exon 3.

ACCESSION U83428

VERSION U83428.1 GI:2688808

KEYWORDS 3 of 9

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 417)

AUTHORS Martin, M.G., Gutierrez, E.M., Lam, J.T., Li, T.W. and Wang, J.

TITLE Genomic cloning and structural analysis of the murine polymorphic
receptor (p1gr) gene and promoter region
JOURNAL Gene 201 (1-2), 189-197 (1997)
MEDLINE 98072444
REFERENCE 2 (bases 1 to 417)
AUTHORS Martin.M.G., Gutierrez,E.M., Lam.J.T., Li.T.W.H. and Wang,J.
TITLE Direct Submission
JOURNAL Submitted (31-DEC-1996) Pediatrics, UCLA School of Medicine, 10833
Le Conte Ave, Los Angeles, CA 90095, USA
FEATURES
source Location/Qualifiers
1. .417
/organism="Mus musculus"
/strain="129SVJ"
/db_xref="taxon:10090"
43. .387
/gene="p1gr"
/number=3
exon
BASE COUNT 105 a 120 c 102 g 90 t
ORIGIN

alignment_scores:
Quality: 188.00 Length: 97
Ratio: 2.648 Gaps: 3
Percent Similarity: 73.196 Percent Identity: 40.206

alignment_block:

US-09-135-238B-2_COPY_18_272 x MMP1GR03 ..

Align seg 1/1 to: MMP1GR03 from: 1 to: 417

```

13 GlyGlySerValThrLeuLysCysProLeuProGlu.....Me 25
|||||
96 GGGGACTCTGTTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 145
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
|||||
146 GCACACCCGGAATACTGTTGCCGACAAAGGAGCC...AGCGGCATGTGCA 192
42 lYThrValValSerThrThrAsnPhelLeLysAlaGluTyrLysGlyArg 58
|||||
193 CAACGCTCATCTCTTCAATGGCTACCTCTCCAGGAGTATTTCAGGCAGA 242
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||||
243 GCCAACCTCATCACTCCAGAGACACACACATTTGTGATTAACATTGA 292
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
|||||
293 GCAGCTCACCCAGGACGACACTGGGAGCTACAAGTGTGGCTGGGT...A 339
92 snThrAspArgGlyThrGlnLysValThrLeuAsnVal 105
|||||
340 CCAGTAACCGAGGCGTGTCTTCGATGTCAAGCTGCAGCTGGAGGTC 380

```

seq_name: gb_ro:MMP1GR3

seq_documentation_block:
LOCUS MMP1GR3 505 bp DNA ROD 09-APR-1998
DEFINITION Mus musculus p1gr gene, exon 3.
ACCESSION Y16525
VERSION Y16525.1 GI:3046750
KEYWORDS p1gr gene; polymorphic immunoglobulin receptor.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 505)
AUTHORS de Groot,N., Vollebregt,E., Lee,S.H., Verbeet,M.P. and de Boer,H.A.
TITLE Molecular cloning and exon-intron organization of the gene encoding the murine polymorphic immunoglobulin receptor
JOURNAL unpublished
REFERENCE 2 (bases 1 to 505)

AUTHORS de Groot,N.
TITLE Direct Submission
JOURNAL Submitted (11-FEB-1998) N. De Groot, Medical Biotechnology Dept.,
Leiden University, P.O. box 9502, 2300 RA Leiden, NETHERLANDS
FEATURES
source Location/Qualifiers
1. .505
/organism="Mus musculus"
/strain="129"
/db_xref="taxon:10090"
/tissue_type="liver"
77. .421
/gene="p1gr"
77. .421
/gene="p1gr"
/label=ex3
/usedin=Y16524:p1gr_cds
/usedin=Y16524:p1gr_mrna
BASE COUNT 130 a 136 c 127 g 112 t
ORIGIN

alignment_scores:
Quality: 188.00 Length: 97
Ratio: 2.648 Gaps: 3
Percent Similarity: 73.196 Percent Identity: 40.206

alignment_block:

US-09-135-238B-2_COPY_18_272 x MMP1GR3 ..

Align seg 1/1 to: MMP1GR3 from: 1 to: 505

```

13 GlyGlySerValThrLeuLysCysProLeuProGlu.....Me 25
|||||
130 GGGGACTCTGTTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 179
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
|||||
180 GCACACCCGGAATACTGTTGCCGACAAAGGAGCC...AGCGGCATGTGCA 226
42 lYThrValValSerThrThrAsnPhelLeLysAlaGluTyrLysGlyArg 58
|||||
227 CAACGCTCATCTCTTCAATGGCTACCTCTCCAGGAGTATTTCAGGCAGA 276
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||||
277 GCCAACCTCATCACTCCAGAGACACACACATTTGTGATTAACATTGA 326
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
|||||
327 GCAGCTCACCCAGGACGACACTGGGAGCTACAAGTGTGGCTGGGT...A 373
92 snThrAspArgGlyThrGlnLysValThrLeuAsnVal 105
|||||
374 CCAGTAACCGAGGCGTGTCTTCGATGTCAAGCTGCAGCTGGAGGTC 414

```

seq_name: gb_pat:E15253

seq_documentation_block:
LOCUS E15253 29392 bp DNA PAT 28-JUL-1999
DEFINITION Mouse p1g receptor gene.
ACCESSION E15253
VERSION E15253.1 GI:5709936
KEYWORDS JP 1998057066-A/1.
SOURCE Mus sp.
ORGANISM Mus sp.

REFERENCE 1 (bases 1 to 29392)
AUTHORS Sato,T. and Kushi,A.
TITLE MURINE PIG RECEPTOR PROTEIN GENE
JOURNAL Patent: JP 1998057066-A 03-MAR-1998;
YAKULT HONSHA CO LTD
COMMENT OS Mus sp. (mouse)

PN JP 1998057066-A/1
PD 03-MAR-1998
PI 19-AUG-1996 JP 1996217154
PI SATO TAKASHI, KUSHIRO AKIRA
PC C12N15/09,C07H21/04,C07K14/705;
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FH source 1..29392
FT /organism="Mus sp."
FT /clone_lib="genomic DNA from female 129SVJ FT
FT mouse liver in
FT lambda Fix vector'.
FT Location/Qualifiers
1..29392
/organism="Mus sp."
/db_xref="taxon:10095"
BASE COUNT 8318 a 6747 c 6514 g 7813 t
ORIGIN
alignment_scores:
Quality: 188.00 Length: 97
Ratio: 2.648 Gaps: 3
Percent Similarity: 73.196 Percent Identity: 40.206
alignment_block:
US-09-135-238B-2_COPY_18_272 x E15253 ..
Align seg 1/1 to: E15253 from: 1 to: 29392
13 GlyGlySerValThrLeuLysCysProLeuProGlu.....Me 25
||| |||||:|||| ||| |||||:|||||
18245 GCGACTCTGTTCCATCAGTCTACTACCCAGACACCTGTGTCACCG 18294
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
|||||:|||| ||| |||||:|||| ||| |||||:|||||
18295 GCACACCGGAATACTGTGTGCCGACAGGAGCC...AGCGCATGTGCA 18341
42 lyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 58
|||||:||||:||||:||||:||||:||||:||||:||||:||||:|||||
18342 CAAGCTCATCTCTTCAATGGCTACTCTCCAAGGAGTATTACGCGAGA 18391
59 ValThrLeuLysGlyTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||||:||||:||||:||||:||||:||||:||||:||||:||||:|||||
18392 GCCAACCTCATCAACTTCCAGACACACACATTTGTGATTAACTTGA 18441
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
|||||:||||:||||:||||:||||:||||:||||:||||:||||:|||||
18442 GCAGCTCACCCAGGAGCAGACTGGGAGCTACAAAGTGTGCGCTGGGT...A 18498
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 105
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18489 CCAGTAACCGAGCGCTGCTCTTCGATGTCAGCCTGGAGGTC 18529
seq_name: gb_ro:AB001489
seq_documentation_block:
LOCUS AB001489 29392 bp DNA 14-APR-2000
DEFINITION Mus musculus DNA for polyimmunoglobulin receptor, complete cds.
ACCESSION AB001489
VERSION AB001489.1 GI:2804245
KEYWORDS polyimmunoglobulin receptor.
SOURCE Mus musculus (strain:129SVJ) female liver DNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Kushiro,A. and Sato,T.
REFERENCE
1 (sites)
AUTHORS
TITLE
POLYMERIC IMMUNOGLOBULIN RECEPTOR GENE OF MOUSE: SEQUENCE,
STRUCTURE AND CHROMOSOMAL LOCATION
JOURNAL
Gene 204 (1-2), 277-282 (1997)

MEDLINE 98094283
REFERENCE 2 (bases 1 to 29392)
AUTHORS Sato,T.
TITLE Direct Submission
JOURNAL Submitted (26-FEB-1997) to the DDBJ/EMBL/GenBank databases. Takashi
Sato, Yakult Central Institute for Microbiological Research,
Culture Collection Center; 1796 Yaho, Kunitachi-shi, Tokyo 186,
Japan (Tel:425-77-8960, Fax:425-77-3020)
Location/Qualifiers
1..29392
/organism="Mus musculus"
/strain="129SVJ"
/db_xref="taxon:10090"
/chromosome="1"
/map="1E3"
/sex="female"
/tissue_type="liver"
CAAT_signal 2470..2474
TATA_signal 2580..2585
CAAT_signal 2878..2882
CAAT_signal 3176..3180
CAAT_signal 3251..3255
TATA_signal 3350..3355
5'UTR 3389..3501
/note="based on the putative transcription start"
3389..3501
/number=1
3502..11067
/number=1
11124..27591
/gene="pigr"
join(11124..11166,18192..18536,21130..21783,23022..23354,
23908..24240,25133..25325,25686..25813,26183..26314,
26426..26484,27496..27591)
/gene="pigr"
/function="mediates secretion of IgA and IgM"
/codon_start=1
/product="polyimmunoglobulin receptor"
/protein_id="BAA24431.1"
/db_xref="GI:2804246"
translation="MRLYLFTLLVTVFSGVSTKSPFGPQVSVSTEGDSVSTICVYPD
TSVNRHTRKYWCRCQAGSMCTTLISSNGYLSKEYSGRANLINFENNTFVINIQLTQ
DDTGSYKGLGTSNRGLSFDYSLOYPELPSDRHYTKDIGNVTIECPFKRENAP
SKKSLCKNTNSCELYIDSTEKNPSEYIGRAKLFMKGTDLTIVYVNIHLTHNDAGLY
TCDVLIINTLGRDPDFEGRILITPKDDNGRFSVLITGLRKEDAGHYQCGAHSSGLPOE
GWPIQTWLFVNEESTIPNRRSVKVTGGSVAIACYPNPKESSLKYCWKWDGNG
HCPVLVGTQAOVQEEYEGRLALFDQPGNGTIVTILNOLTTEDAGFYCLTNGDSWRT
TIELOVAEATREPNEVTPONATAVLGETFTVSYCHYCKFYQKQYKCKWNSKGCCHIL
PSHDEGARQSSVSCDQSLYSMTLNPVSKDEGWICGVCQQTGYETTAIYIYAVEE
KIRGSSHVNPTDANARAKVALEEVVDSISSEKENKAIPNPGPFANEREIKNVQDQAG
ENRSGDAGSADGSSSSKVLFTSLPLGLVLAVGAIAVYAVRHRKKNVDRMSIS
SYRTDISMADFKNSRDLGNDNMGASPDQTQOTVIEGKDEIVTTTCTAEPEESKAKR
SSKEADMAYSAFLLQSSSTIAAQVHDGPOEA"
11124..11166
/gene="pigr"
/number=2
11167..18191
/gene="pigr"
/number=2
18192..18536
/gene="pigr"
/number=3
18537..21129
/gene="pigr"
/number=3
21130..21783
/gene="pigr"
/number=4
21784..23021
/gene="pigr"
/number=4
23022..23354
exon
intron
exon
intron
exon
intron
exon

```
18442 GCAGCTCACCCAGGACGACTGGGAGCTACAAGTGTGGCTGGGT...A 18488
92 snThrasArgGlyLysThrGlnLysValThrLeuAsnVal 105
   :::::::::::::: :::: ||::::::::::|||
18489 CCAATAACGAGCGCTGCTTCGTGATGTCAGCGCTGAGGTC 18529

seq_name: gb_ro:RNIGR

seq_documentation_block:
LOCUS      RNIGR               3269 bp      mRNA           ROD      22-MAR-1995
DEFINITION Rat mRNA for polymeric immunoglobulin receptor.
ACCESSION  X15741
VERSION    X15741.1 GI:56464
KEYWORDS   immunoglobulin receptor; polymeric immunoglobulin receptor;
           transmembrane protein.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE  1 (bases 1 to 3269)
AUTHORS   Banting,G.
TITLE     Direct Submission
JOURNAL   Submitted (05-JUL-1989) Banting G., EMBL, Meyerhofstr 1, 6900
           Heidelberg
REFERENCE  2 (bases 1 to 3269)
AUTHORS   Banting,G., Brake,B., Braghetta,P., Luzio,J.P. and Stanley,K.K.
JOURNAL   Unpublished
FEATURES   Location/Qualifiers
            source
              1..3269
              /organism="Rattus norvegicus"
              /db_xref="taxon:10116"
              /tissue_type="liver"
              /clone_lib="cDNA, TRL1"
              /clone="GORF1,8,43,4339"
sig_peptide
              74..127
              /product="rat polymeric immunoglobulin receptor signal
              peptide (AA -18 to -1)"
              74..2383
CDS        /note="rat polymeric immunoglobulin receptor (AA -18 to
              751)"
              /codon_start=1
              /protein_id="CAA33758.1"
              /db_xref="GI:56465"
              /db_xref="SWISS-PROT:P15083"
              /translation="MRLSIFALLVTFVSGVSTQSPFGPQDVSSIEGNSVSTCYVPD
              TSVNRHTRKYWCQNGCYCATLISSNGYLSKEYSGRASLNFNPENSTFVINTAHLTQ
              EDTGYKGLGTFRNGLFVDSVSVQYPPFPNDPHVTKDIGFTVTECRKEGNAH
              SKKSLCKRGEACEVVIDSTEYVDPSTYKDRALLFMKGSRLDFYNISHLIFSDAGLY
              VCQAGGFSADKNNADLVLEPEPELLYKDLRSSVTFECDLGREVANDAKYLCKRANKE
              TCDVIINTLGRDPAFEGRIILLTPRDDNGRFSVLITGLRKEDAGHYOCGAHSSGLPQE
              GMPVQALQVOEGYEGRLALFDQPGSGAYTVLNQLTQDSGFYWCITFDGDSWRMT
              RCPVLVGTALVQEGYEGRLALFDQPGSGAYTVLNQLTQDSGFYWCITFDGDSWRMT
              TTELOVAETKPDLEVPONATAVIGETFTISCHVPCKFYSEKYCKWSNDGCHIL
              PSHDEGARQSSVSCDSSQIVSMTLNPKVEDGHWGKVGEGOVGYCTTAIYVAVEE
              RYRSPHINFDANARAKDAPEEAMESSYVREDENKANLDPLEADEREIQNAGDAQ
              ENRASGNAGSAGGSGSSKVLFTSLVPLGLVLAGVAYAVWVARYVHRKNVDKMSISSY
              RTDISMGDFRNSRDLGGNDMGATPDQETVLEGGKDEIETTECTTEPEESKKAARSS
              128..2380
              /product="rat polymeric immunoglobulin receptor mature
              peptide (AA 1 to 751)"
mat_peptide
              890 a      831 c      850 g      698 t
BASE COUNT
ORIGIN

alignment_scores:
Quality: 185.50      Length: 164
Ratio: 1.855      Gaps: 6
Percent Similarity: 60.976      Percent Identity: 30.488

alignment_block:
US-09-135-238B-2_COPY_18_272 x RNIGR
```

```
/gene="pIgr"
/number=5
23355..23907
/gene="pIgr"
/number=5
23908..24240
/gene="pIgr"
/number=6
24241..25132
/gene="pIgr"
/number=6
25133..25325
/gene="pIgr"
/number=7
25326..25685
/gene="pIgr"
/number=7
25686..25913
/gene="pIgr"
/number=8
25914..26182
/gene="pIgr"
/number=8
26183..26314
/gene="pIgr"
/number=9
26315..26425
/gene="pIgr"
/number=9
26426..26484
/gene="pIgr"
/number=10
26485..27495
/gene="pIgr"
/number=10
27496..29008
/number=11
27592..29008
/note="based on the polyadenylation site"
28983..28988
polyA_signal
29008
BASE COUNT      8318 a      6747 c      6514 g      7813 t
ORIGIN

alignment_scores:
Quality: 188.00      Length: 97
Ratio: 2.648      Gaps: 3
Percent Similarity: 73.196      Percent Identity: 40.206

alignment_block:
US-09-135-238B-2_COPY_18_272 x AB001489 ..
Align seg 1/1 to: AB001489 from: 1 to: 29392
13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
||| |||||:||||| ||| |||||
18245 GCGACTCTGTTCATCATCGTCTACTACCCAGACACCTGTCAACCG 18294
25 ThrValArgIleThrLeuLysCysArgGluMetAlaGlySerGlyThrCysG 42
||||:||||| ||| |||||:||||| ||| ||||| |||
18295 GCACACCCGGAATATGCTGCCGACAGGAGCC...AGCGGATGTGCA 18341
42 IyThrValValSerThrThrAsnPheIleLysAlaGluThrLysGlyArg 58
||||:|||||:|||||:|||||:|||||:|||||:|||||
18342 CAACGCTCATCTCTCAATGGCTACCTTCCAAGGAGTATTCAGGCAGA 18391
59 ValThrLeuLysGlnThrProArgLysAsnLeuPheLeuValGluValThr 75
||||:||||| :||||:||||| ||| |||||:|||||
18392 GCCAACCTCATCACTCCGAGAACCAACACATTTGTGATTACATTGA 18441
75 rGlnLeuThrGluSerAspSerGlyValThrAlaCysGlyAlaGlyMetA 92
|||||:|||||:|||||:||||| ||| ||||| ||| |||
```

Align seg 1/1 to: RNIGR from: 1 to: 3269

```

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
||||:||||:||||:||||:||||:||||:||||:||||:||||:
170 GGTAAGTGGTCTCCATCAGTGGCTACTACCCAGACACCTGTGCAACGG 219
||||:||||:||||:||||:||||:||||:||||:||||:||||:
25 ThisValArgIleThrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||||:||||:||||:||||:||||:||||:||||:||||:||||:
220 GCACACCCGGAATACTGGTGGCCGACAGGAGCC...AACGGCTACTGG 266
||||:||||:||||:||||:||||:||||:||||:||||:||||:
42 lyThrValValSerThrThrAsnPhelleLysAlaGluThrLysGlyArg 58
||||:||||:||||:||||:||||:||||:||||:||||:||||:
267 CAACCTCATCTCTTCAATGGCTACTCTCGAAGGAGTATTTCAGGCAGA 316
||||:||||:||||:||||:||||:||||:||||:||||:||||:
59 ValThrLeuLysGlnThrProArgLysAsnLeuPheLeuValGluValTh 75
||||:||||:||||:||||:||||:||||:||||:||||:||||:
317 GCCACCTCATCACTCCAGAGAAATAGCATTGTGATTAAACATTGC 366
||||:||||:||||:||||:||||:||||:||||:||||:||||:
75 rGlnLeuThrGluSerAspSerGlyValThrAlaCysGlyAlaGlyMetA 92
||||:||||:||||:||||:||||:||||:||||:||||:||||:
367 ACATCTCACCCAGGAGGACACTGGGAGCTACAAGTGTGGTCTGGT...A 413
||||:||||:||||:||||:||||:||||:||||:||||:||||:
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
||||:||||:||||:||||:||||:||||:||||:||||:||||:
414 CCACCTAACCGAGGCTGTTTTCGATGTGCAGCTGGAGTC..... 454
||||:||||:||||:||||:||||:||||:||||:||||:||||:
109 TyrGluProSerTrpGluGluGlnProMetProGluThrProLysTrpPh 125
||||:||||:||||:||||:||||:||||:||||:||||:||||:
455 .....AGCCAGGTCTCTGAGTTCCTCCCAAAATGACAC 483
||||:||||:||||:||||:||||:||||:||||:||||:||||:
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerLysP 142
||||:||||:||||:||||:||||:||||:||||:||||:||||:
484 CCATGTC.....TACACAAAGG 500
||||:||||:||||:||||:||||:||||:||||:||||:||||:
142 heValThrArgValThrThrProAlaGlnArgGlyLysValProVal 158
||||:||||:||||:||||:||||:||||:||||:||||:||||:
501 ACATAGGAGAGAACTGTGACCACTGGAATCCGCTTTCAGAGAGGGAATGCT 550
||||:||||:||||:||||:||||:||||:||||:||||:||||:
159 HisHisSer.SerProThrThrGlnIleThrHisArgPro 171
||||:||||:||||:||||:||||:||||:||||:||||:||||:
551 CATAGCAAGAAATCCCTGTGTAAGAAGAGAGAGAGGCCT 590
||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

seq_name: gb_om:AB032195

```

seq_documentation_block: 2461 bp mRNA MAM 03-FEB-2000
LOCUS AB032195
DEFINITION Sus scrofa mRNA for poly-Ig receptor, complete cds.
ACCESSION AB032195
VERSION AB032195.2 GI:6863079
KEYWORDS poly-Ig receptor.
SOURCE Sus scrofa female cDNA to mRNA, clone:SC1.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
Sone,T. and Kumura,H.
Sone,T. and Kumura,H.
TITLE Porcine mammary gland cDNA clone, similar to poly-Ig receptor
JOURNAL Published Only in DataBase (1999) In press
REFERENCE 2 (bases 1 to 2461)
AUTHORS Sone,T. and Kumura,H.
DIRECT SUBMISSION
TITLE Direct Submission
JOURNAL Submitted (07-SEP-1999) to the DDBJ/EMBL/GenBank databases. Teruo
Sone, Hokkaido University, Laboratory of Applied Microbiology,
Faculty of Agriculture, Kita-9, Nishi-9, Kita-ku, Sapporo, Hokkaido
060-8589, Japan (E-mail:sonet@chem.agr.hokudai.ac.jp,
Tel:81-11-706-2502, Fax:81-11-706-4961)
On Feb 3, 2000 this sequence version replaced gi:5881588.
Sequence updated (01-Feb-2000).
COMMENT
FEATURES
Location/Qualifiers
source
1..2461
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone="SC1"

```

```

/sex="female"
58..111
/product="putative"
58..2334
/codon_start=1
/product="poly-Ig receptor"
/protein_id="BAA84283.2"
/db_xref="GI:6863080"
/translaton="MTRFFYACLLAIFPVVSMKSPIFGPODVSSVEGSSVSIRCYTPA
TSVNRHSRYWCRIGAGKRCCTTLISSEGYISKDYKGRANLTFNPNENGTFYMDIGHLTR
GDSGLYKCGGISRGSLGSLFVLEVSQPGQIGDVHVTYADLGSVTYINCPFKSENAQ
PKSVYKIKGOIRVLVDSNGLNNTNRAHLSIOGTNOLVFSVFNRIQLRDAGTY
ICQAGDESSADLOVLKPEPELYGDLRGSVTFDCALGOEMANYAFELCOLKNGKTCN
VYNTLIGKKAQDEPEGRILLTPKNSHFSVHITLKRDEAGHLCGHPDPEGEGMPV
QAWQLFINEDTIMIPRSPVVGVSVAVTCYPNKRNKSLKRWENENGRCPQ
LVESGLVNEQYERLALYBEPNGTFTVILNQLNRDAGFYWCLNEDSRWSTVEL
KIVGEPNLKLPENVTAWGETLKSCHFPCKFYQKYWKCKNSNTGCRALPQDQEQ
SOAFVNCDDKKSQIISLNLNPKRDEGWYKCGVDGLHYGETGAVTVAVPQKAKGSD
ARLANAPAPAEDAIEPRARETENEVLDPSLIAEDRAVKDGGSPASGSGVPAPOSS
VCGGSGSKAVSTLPLALVNLVAVNTIGVLRHHRKNVDRIISIRSTRIDISDFES
SRDGAHDNAGATLDTOETSLGGKDRFTTTTENTMEDKEPKRAKRSKEADKAFITF
LQANSIIAATQNGPREA"
BASE COUNT 608 a 699 c 708 g 446 t
ORIGIN

```

```

alignment_scores:
Quality: 181.00 Length: 118
Ratio: 2.207 Gaps: 4
Percent Similarity: 69.492 Percent Identity: 33.898
alignment_block:
US-09-135-238B-2_COPY_18_272 x AB032195 ..
Align seg 1/1 to: AB032195 from: 1 to: 2461

```

```

2 IleLeuProGluValLysValGluGlyGluLeu..... 12
||||:||||:||||:||||:||||:||||:||||:||||:||||:
91 ATCTTCCCGAGTGGTCTCCATGAAGAGTCCCATATTCGTGTCCTCCAGGATGT 140
||||:||||:||||:||||:||||:||||:||||:||||:||||:
13 .....GlyGlySerValThrIleLysCysProLeuPro.... 23
||||:||||:||||:||||:||||:||||:||||:||||:||||:
141 GAGCAGCGTGGAGGACGCTCGGTGTCCATCAGATGCTACTACCCAGCCA 190
||||:||||:||||:||||:||||:||||:||||:||||:||||:
24 .....GluMethHisValArgIleThrLeuCysArgGluMetAlaGly 37
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 CTTCCGTCACCGCATTCCTCGGAAGTACTGTGTCGCA...ATAGGAGCC 237
||||:||||:||||:||||:||||:||||:||||:||||:||||:
38 SerGlyThrCysGlyThrValValSerThrThrAsnPhelleLysAlaG1 54
||||:||||:||||:||||:||||:||||:||||:||||:||||:
238 AAGGCCCTGCACACCCCTCATCTCTCGGAGGCTACATCTCCAAGGA 287
||||:||||:||||:||||:||||:||||:||||:||||:||||:
54 uTyrLysGlyArgValThrLeuLysGlnThrProArgLysAsnLeuPheL 71
||||:||||:||||:||||:||||:||||:||||:||||:||||:
288 CTACAAGGCGAGAGCAACCTCACCACTCCAGAGAACGCGACCTTCG 337
||||:||||:||||:||||:||||:||||:||||:||||:||||:
71 euValGluValThrGlnLeuThrGluSerAspSerGlyValThrAlaCys 87
||||:||||:||||:||||:||||:||||:||||:||||:||||:
338 TGATGGACATTGGCCACCTACCGCGGTGACTCTGTGGGCTCTACAAGTGT 387
||||:||||:||||:||||:||||:||||:||||:||||:||||:
88 GlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAs 104
||||:||||:||||:||||:||||:||||:||||:||||:||||:
388 GGTCTGGGCATTAGCAGC...CGAGGCCTGTCTTTTGTGATGTGAGCCTGA 434
||||:||||:||||:||||:||||:||||:||||:||||:||||:
104 nVal 105
||||:
435 GGTC 438

```

seq_name: gb_pr2:S4343503

```

seq_documentation_block:
LOCUS S4343503 345 bp DNA PRI 08-MAY-1993
DEFINITION transmembrane secretory component [human, leukocytes, Genomic, 345

```

[illegible]

us-09-135-238b-2_copy_18_272.p2n.rge

us-09-135-238b-2_copy_18_272.p2n.rge

OM of: US-09-135-238B-2_COPY_18_272 to: Issued_Patents_NA.* out_format : pfs
Date: Sep 12, 2000 6:53 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=framet_p2n.model -DRV=xlp
-O/cgn2_1/USPTO_spool/US09135238/runat_24082000_091324_17895/app_query.fasta_1.1282
-DB=Issued_Patents_NA -OFMT=fastap -SUFFIX=p2n.rni
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOAPEXT=0.000 -CGAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FCGPOP=6.000 -FCGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCLALIGN=200
-THRCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000
-USER=US09135238_cgn1_1_68 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-135-238B-2_COPY_18_272
Query length: 255
Database: Issued_Patents_NA.*
Database sequences: 243080
Database length: 68777915
Search time (sec): 93.130000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-7	+	197.00	340.00	1.9e-11	3095	!
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-9	+	185.50	317.08	3.6e-10	3269	!
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-3	+	180.50	308.80	1.0e-09	2919	!
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-5	+	179.50	304.19	1.9e-09	3630	!
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-642-406A-21	+	166.50	287.15	1.7e-08	1875	!
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-642-406A-20	+	166.50	279.42	4.5e-08	3517	!
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-1	+	166.50	279.42	4.5e-08	3517	!
/cgn2_6/ptodata/1/ina/5A/COMB.seq:US-08-286-740-3	+	102.00	146.97	1.07	6557	!
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-09576-3	+	102.00	146.97	1.07	6557	!
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-378-939-9	+	100.50	161.27	0.1716	1617	!
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-887-352B-1	+	99.50	142.97	1.79	6127	!
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-935-450-10	+	99.00	165.38	0.1013	914	!
/cgn2_6/ptodata/1/ina/5A/COMB.seq:US-07-841-646-3	-	97.00	125.30	17.29	17410	!
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-147-023-3	-	97.00	125.30	17.29	17410	!
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-447-570-3	-	97.00	125.30	17.29	17410	!
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-449-700-3	-	97.00	125.30	17.29	17410	!
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-449-699A-3	-	97.00	125.30	17.29	17410	!
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-157-101A-9	+	96.50	155.75	0.3484	1350	!
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-157-101A-6	+	96.50	153.85	0.4446	1576	!
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-932-376A-1	+	93.50	146.84	1.09	1738	!
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-463-587A-24	+	92.00	141.17	2.26	2178	!
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-463-667A-2	+	92.00	141.17	2.26	2178	!
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-923-854-24	+	92.00	141.17	2.26	2178	!
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US91-09133-25	+	92.00	141.17	2.26	2178	!
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-149-099C-3	+	91.50	122.49	24.82	9209	!
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-476-275-2	+	91.50	122.49	24.82	9209	!
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-478-967A-3	+	91.50	122.49	24.82	9209	!
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-819-866-2	+	91.50	113.60	77.61	18986	!
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-09-023-715-2	+	91.50	113.60	77.61	18986	!
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-592-126-75	-	91.00	147.81	0.9637	1083	!
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-932-376A-3	+	91.00	139.00	2.99	2220	!
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-09-027-449-59	+	90.50	148.76	0.8539	927	!
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-09-027-449-61	+	90.50	124.71	18.65	6563	!
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-09-027-449-68	+	90.50	122.10	26.09	8120	!
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-09-027-449-68	+	90.50	119.57	36.06	9972	!
/cgn2_6/ptodata/1/ina/5A/COMB.seq:US-08-276-099A-1	+	90.00	133.18	6.30	3046	!
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-781-890-1	+	90.00	133.18	6.30	3046	!
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-539-205A-5	+	90.00	132.47	6.90	3226	!
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-860-882A-28	+	89.50	148.99	0.8287	777	!
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-394-326-1	+	89.00	135.99	4.39	2070	!

/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-00031-18 + 88.50 135.89 4.45
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-00031-20 + 88.50 135.87 4.46
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-00031-10 + 88.50 135.81 4.49
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-00031-22 + 88.50 135.81 4.49
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-00031-12 + 88.50 134.24 5.49
seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-7
seq_documentation_block:
; Sequence 7, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 1
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3095 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; TOPOLOGY: DESCRIPTION: Mouse Polyimmunoglobulin Receptor
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 85....2400
; US-08-434-000A-7
alignment_scores:
Quality: 197.00 Length: 148
Ratio: 2.165 Gaps: 5
Percent Similarity: 61.486 Percent Identity: 32.432
alignment_block:
US-09-135-238B-2_COPY_18_272 x US-08-434-000A-7 ..
Align seq 1/1 to: US-08-434-000A-7 from: 1 to: 3095
13 GlycylSerValThrIleLysCysProLeuProGlu.....Me 25

```
181 GGGGACTCTGTTTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 230
||||| ||||| ||||| ||| |||||
25 tHisValArgIleThrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||||| ||| ||||| ||| ||||| |||
231 GCACACCCGGAATACTGTCGCCAAGAGGCC...AGCGCATGTGCA 277
||||| ||||| ||||| ||||| |||||
42 lyThrValValSerThrThrAsnPheIleLysAlaGluThrLysGlyArg 58
||||| ||||| ||||| ||||| |||||
278 CAACGCTCATCTCTCAATGCTACTCTCCAAGGAGTATTTCAGGCAGA 327
||||| ||||| ||||| ||||| |||||
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
||||| ||||| ||||| ||||| |||||
328 GCAACCTCATCACTTCCAGAGAACACACATTTGTGATTAAACATTGA 377
||||| ||||| ||||| ||||| |||||
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
||||| ||||| ||||| ||||| |||||
378 GCAGCTCACCCAGCAGACACTGGGAGCTACAAGTGTGGCTGGGT...A 424
||||| ||||| ||||| ||||| |||||
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
||||| ||||| ||||| ||||| |||||
425 CCAGTAACCGAGGCTGCTCTCGATGTCAGCCTGGAGGTC..... 465
||||| ||||| ||||| ||||| |||||
109 TyrGluProSerTrpGluGluGlnProMetProGluThrProLysTrpPh 125
||||| ||||| ||||| ||||| |||||
466 .....AGCCAGGCTCTGAGTGCAGGAGTGACAC 494
||||| ||||| ||||| ||||| |||||
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerLysP 142
||||| ||||| ||||| ||||| |||||
495 CCAGCTC.....TACACAAAGGACATAGGCA 520
||||| ||||| ||||| ||||| |||||
142 heValThrArgValThrProAlaGlnArgGlyLysValPro 156
||||| ||||| ||||| ||||| |||||
521 GAATGTGACCATGATGCCTTCAAAAGGGAGATGTCCC 564
||||| ||||| ||||| ||||| |||||
seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-9
seq_documentation_block:
; Sequence 9, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434, 000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; TELETYPE: SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FUNCTION: DESCRIPTION: Rat Polyimmunoglobulin Receptor
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 74....2383
; US-08-434-000A-9

alignment_scores:
Quality: 185.50 Length: 164
Ratio: 1.855 Gaps: 6
Percent Similarity: 60.976 Percent Identity: 30.488

alignment_block:
US-09-135-238B-2_COPY_18_272 x US-08-434-000A-9 ..
Align seg 1/1 to: US-08-434-000A-9 from: 1 to: 3269
13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
||||| ||||| ||||| ||| |||||
170 GGTAACTCGGTCTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 219
||||| ||||| ||||| ||||| |||||
25 tHisValArgIleThrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||||| ||||| ||||| ||||| |||||
220 GCACACCCGGAATACTGTCGCCAAGAGGCC...AACGCTACTGCG 266
||||| ||||| ||||| ||||| |||||
42 lyThrValValSerThrThrAsnPheIleLysAlaGluThrLysGlyArg 58
||||| ||||| ||||| ||||| |||||
267 CAACCTCATCTCTCAATGCTACTCTCGAAGGAGTATTTCAGGCAGA 316
||||| ||||| ||||| ||||| |||||
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
||||| ||||| ||||| ||||| |||||
317 GCAGCTCATCACTTCCAGAGATAGCACATTTGTGATTACATTGC 366
||||| ||||| ||||| ||||| |||||
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
||||| ||||| ||||| ||||| |||||
367 ACATCTCACCCAGGAGACACTGGGAGCTACAAGTGTGCTGGGT...A 413
||||| ||||| ||||| ||||| |||||
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
||||| ||||| ||||| ||||| |||||
414 CCAGTACCCGAGGCTGTTTTCGATGTCAGCTGGAGGTC..... 454
||||| ||||| ||||| ||||| |||||
109 TyrGluProSerTrpGluGluGlnProMetProGluThrProLysTrpPh 125
||||| ||||| ||||| ||||| |||||
455 .....AGCCAGGTTCTGAGTTCACAAAGAGGGAATGCT 483
||||| ||||| ||||| ||||| |||||
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerLysP 142
||||| ||||| ||||| ||||| |||||
484 CCATGTC.....TACACAAAGG 500
||||| ||||| ||||| ||||| |||||
142 heValThrArgValThrProAlaGlnArgGlyLysValProProVal 158
||||| ||||| ||||| ||||| |||||
501 ACATAGCAGAACTGTGACCATCGAATGCCGTTTCAAGAGGGAATGCT 550
||||| ||||| ||||| ||||| |||||
159 HisHisSer_SerProThrThrGlnIleThrHisArgPro 171
||||| ||||| ||||| ||||| |||||
551 CATACCAAGAAATCCCTGTGTAAGAGAGAGAGAGGCT 590
||||| ||||| ||||| ||||| |||||
seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-3
seq_documentation_block:
; Sequence 3, Application US/08434000A
; Patent No. 6046037
```

```
;
;
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; COUNTRY: California
; COUNTRY: U.S.A.
; ZIP: 90071
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 1
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2919 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; TOPOLOGY: DESCRIPTION: Human polyimmunoglobulin Receptor
;
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 235....2472
;
; US-08-434-000A-3
;
; alignment_scores:
; Quality: 180.50 Length: 106
; Ratio: 2.314 Gaps: 4
; Percent Similarity: 73.585 Percent Identity: 38.679
;
; alignment_block:
; US-09-135-238b-2_COPY_18_272 x US-08-434-000A-3 ..
;
; Align seg 1/1 to: US-08-434-000A-3 from: 1 to: 2919
;
; 4 ProgluValLysValGluGlyGluLeuLeuGlyGlySerValThrLeuLysCys 20
; ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
; 253 CCGAG...GAGTGAATAGTGGAGGTAACCTACGTCTCCATCAGCTG 299
; | ||| :|||:|||||:|||||:|||||:|||||:|||||:
; 20 sProLeuPro.....GluMethHisValArgLeuTyrLeuCysA 33
; | ||| :|||:|||||:|||||:|||||:|||||:|||||:
; 300 CTACTACCCACCCACCTCTGTCAACCGGCACACCCGGGAAGTACTGGTGCC 349
; ||||| ||| :|||:|||||:|||||:|||||:|||||:|||||:
; 33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
; ||||| ||| :|||:|||||:|||||:|||||:|||||:|||||:
; 350 GGCAGGGAGCTAGAGTGTC...TGCATAACCCCTCATCTCTCGGAGGC 396
```



```

; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,406A
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/591,823
; FILING DATE: 02-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/427,765
; FILING DATE: 27-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Logan, April C.
; REGISTRATION NUMBER: 33,950
; REFERENCE/DOCKET NUMBER: 184.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-642-406A-20

alignment_scores:
  Quality: 166.50      Length: 103
  Ratio: 2.602        Gaps: 4
  Percent Similarity: 62.136      Percent Identity: 42.718

alignment_block:
  US-09-135-238B-2_COPY_18_272 x US-08-642-406A-20  ..
  Align seg 1/1 to: US-08-642-406A-20 from: 1 to: 3517

13 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 25
||| |||||:|||| ||| |||
238 GCGACTCGGTGCCATCATCTACTACCAACAACCTCGTCACCCG 287

25 ThisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||| ||| :||| ||||| ||| ||||| |||
288 GCACACCGGAGTCTGTGTCGGGAA...GAGGAGCGCGCGCTCGC 334
|||:||||| |||:||||| :|||:|||||
335 TGACGCTTGCCTCG...ACCGCTACAGCTCCCGAGTAATCTCCGGAGA 381

59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||:||||| |||:||||| |||:||||| |||
382 GCAAGCTCACCGACTCCCTCTATAAAGGGAGTTTGTGTACTGTGA 431

75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92
|||||:|||||:|||||:||||| ||| |||||:|||||:|||||
432 CCAACTACCCAGCAGCTACGAGGACTACCAAGTGTGGCGTGGAGTCA 481

92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
|| |||||:||||| |||:|||||
482 AC...GGCGGTGGCTGGACTTCGGTGTCTCAAGCTGCTGGTCAGCCAGAG 528

109 TyrGluPro lll
```

```

529 CCAGAGCGCT 537
|||||
seq.name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-1

seq_documentation_block:
; Sequence 1, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 1
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; TELEX: SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; TOPOLOGY: DESCRIPTION: Rabbit polyimmunoglobulin receptor
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 124....2445
; US-08-434-000A-1

alignment_scores:
  Quality: 166.50      Length: 103
  Ratio: 2.602        Gaps: 4
  Percent Similarity: 62.136      Percent Identity: 42.718

alignment_block:
  US-09-135-238B-2_COPY_18_272 x US-08-434-000A-1  ..
  Align seg 1/1 to: US-08-434-000A-1 from: 1 to: 3517

13 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 25
||| |||||:|||| ||| |||
238 GCGACTCGGTGCCATCATCTACTACCAACAACCTCGTCACCCG 287
```

```
25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42 *
   ||| ||| ::| ||||| ||| ||| |||
288 GCACAGCGGGAAGTCTGTGCGGGAA...GAGGAGAGCGCGCGCTGCG 334
42 lYThrValValSerThrThrAsnPhelIleLysAlaGluTyrLysGlyArg 58
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
335 TGACGCTTGCCTCG...ACCGGCTACAGCTCCAGGAATACTCCGGGAGA 381
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 GCAAGCTCACCGACTCCCTGATAAAGGGGAGTTTGTGCTGACTGTGA 431
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
432 CCAACTCACCCAGAACGACTCAGGGAGCTACAAGTGTGCGTGGGAGTCA 481
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
482 AC...GGCCGTGGCTGGACTTCGGTGTCAAGTGTGCTCAGCCAGAAG 528
109 TyrGluPro 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
529 CCAGAGCGCT 537
```

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:us-08-286-740-3

```
seq documentation block:
; Sequence 3, Application US/08286740
; Patent No. 5561053
; GENERAL INFORMATION:
; APPLICANT: Crowley, Craig W.
; TITLE OF INVENTION: METHOD FOR SELECTING HIGH-EXPRESSING
; TITLE OF INVENTION: HOST CELLS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,740
; FILING DATE: 05-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6557 bases
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-286-740-3
```

alignment_scores:
Quality: 102.00
Ratio: 0.843

Length: 252
Gaps: 15

```
Percent Similarity: 48.016 Percent Identity: 24.603
alignment_block:
US-09-135-238B-2_COPY_18_272 x US-08-286-740-3 ..
Align seg 1/1 to: US-08-286-740-3 from: 1 to: 6557

13 GlyGlySerValThrIleLysCysProLeuProGluMethHisValArgIl 29
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1405 GGGGGCTCACTCCGTTTGTGTGTCAGTTTCTGGCTACTCCATCACCTC 1454
29 e.....TyrLeuCysArgGluMetAlaGlySerGly..... 39
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1455 CGGATATAGCTGGAAGTGGATCGCTCAGCGCCCGGTAAGGGCTCGAAT 1504
40 .....ThrCysGlyThrValValSerThrThrAsnPhelIleLysAlaGlu 54
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1505 GGGTTGCATCGATTACGTATGCCGGATCGACTAAGTAT...AACCCCTAGC 1551
55 TyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLe 71
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1552 GTCAGGGCGGTATCACTATATAGTCGGAGATTCCAAAACACATTCATA 1601
71 uValGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysG 88
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1602 CCTGCAGATGAACAGCTCGCTGCTGAGGACACACTGCCGCTCTATTATTGTG 1651
88 lYAlaGlyMetAsn...ThrAspArgGlyLysThrGlnLysValThrLeuAs 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1652 CTCAGGAGCAGCCATATTTCGGCGCTGGCACTTCGCGTGTGGGTCAA 1701
104 nValHisSerGluTyrGluProSerTrpGluGlnPrometPro.... 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1702 G.....GAAACC...TGGTCACCGTCTCCTCGCGCTCCAC 1733
120 .....GluThrProLysTrpPhe.....HisLeuProTyr 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1734 CAAGGGCCCATGGTCTTCCCGCTGGCACCCCTCTCCAGAGACACCTCTG 1783
130 LeuPheGlnMetProAlaTyrAlaSerSerSerLysPheValThrArgVa 146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1784 GGGGCACAGCGCGCTCGCTGCTGCTCA.....AGGAC 1818
146 lThrProAlaGln.....ArgGlyLysValProProValHisH 160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1819 TACTTCCCGGACCGGTCACGGTGTCTGTGAACACTCAGCGCCCTGACCAG 1868
160 lSerSerProThrThrGlnIleThrHisArgProArgValSerArgAla 176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1869 CGGGGTGCACACCTTCGGCGGTCTCTACAGTCTCTCAGGACTCTACTCC 1918
177 SerSer.....ValAlaGlyAspLysProArgThrPh 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1919 TCAGCAGCGGTGGTACTGTGCCCTTAGCAGCTTGGGCACCCAGA..... 1963
187 eleuProSerThrThrAlaSerLysIleSerAlaLeuGluGlyLeuLeuL 204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1964 .....CCTACATCTGCAACGTGAATCAACAAGC..... 1990
204 ysProGlnThrPro.....SerTyrAsnHisHisThr 214
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1991 ..CCAGCAACACAGAGTGGGACAGAAAGTTGAGCCCAATCTTTGTGACA 2038
215 ArgLeu...HisArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyAr 230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2039 AAACCTCACACATGCCACCGGTGCCACGACCTGAACCTCCTGGGGGACCG 2088
230 gGlu 231
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2089 TCAG 2092
```

seq_name: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-09576-3


```

; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1808-118
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 783-6040
; TELEFAX: (202) 783-6031
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 35..92
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 93..1465
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 35..1465
; US-08-378-939-9

```

```

alignment_scores:
  Quality: 100.50      Length: 254
  Ratio: 0.817        Gaps: 11
  Percent Similarity: 48.425  Percent Identity: 24.016

```

```
alignment block:
```

```
US-09-135-238B-2_COPY_18_272 x US-08-378-939-9
```

```
Align seq 1/1 to: US-08-378-939-9 from: 1 to: 1617
```

```

5 GluValLysValGluGlyLeuGlySerValThrIleLysCysPr 21
|||||
119 GAAGTAAG.....AAGCTGGGTCTCGGTGACGGTCTCTCTGCAA 159
21 oLeuProGlu.....MethValArgIleTyrLeuCysArgGluM 35
160 GGCACTGGAGGACCTTCAGCACTATGCTATCAGCTGGTGGCGCAGG 209
35 etAlaGlySerGly.....ThrCysGly 42
|||||
210 CCCTGGACAAGGCTTGAGTGGATGGGAGGATCATCCCTCTTTTGGT 259
43 ThrValValSerThrThrAsnPhelLysAlaGluTyrLysGlyArgVa 59
|||
260 ACACCAACTACTCAGAACTTC.....CAGGGCAGAGT 294
59 lThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValThr 76
|||||
295 CAGGATTACCGGACAAATCCACAGCACAGCCACCATGGAGCTGACTA 344
76 lLeuThrCluSerAspSerGlyValTyrAlaCysGlyAlaGly..... 90
:|||||
345 GCCTGAGATCGAGGACACGGCGGTATCTCTGGGACAGATCGCTAC 394
91 .....MetAsnThrAspArgGlyLysThr.GlnLysValThrLeuAsnV 105
395 AGCAGGCAAAATTTTACCGCGGCCGGTGGCTGGTTCAGCCCTGGGG 444
105 alHisSerCluTyrGluProSerTrpGluGlnProMetProGluThr 121
445 CCAGGGCCACCTTGGTTCACCGTCTCTCAGCCCTCCACCAAGGGCCATCGG 494
122 ProLysTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlase 138
495 TCCTCCCTGGCACCCTCTCTCCAGAGACGACCTCTCTGGGGGCACAGCGCC 544

```

```

138 rSerSerLysPheValThrArgValThrThrProAlaGln.....A 152
:::
545 CTGGCTGGCTGG...TCAAGGACTACTCCCGAACCAGCGGTGACGGTGC 591
152 rGlyLysValProProValHisHisSerSerProThrThrGlnIleThr 168
|||||
592 GTGGAAGTCAAGCGCCCTGACCAAGCGGTGCACACCTTCCCGGTGTC 641
169 HisArgProArgValSerArgAlaSerSerValAlaGlyAspLysProAr 185
:::
642 TACAGTCCCTCAGGACTCTACTCCCTCAGCAGCGGTGTCACCGTCCCT.. 689
185 gThrPheLeuProSerThrThrAlaSerLysLysSerAlaLeuGluGlyL 202
|||||
690 .....CCAGCAGCTTGGCGACCCAGACCTACATCTGCAACGTGAA 729
202 euLeuLysProGlnThrPro.....SerTyrAsnHis 212
:::
730 TCACAAGCCCAACACACCAAGGTGGACAAGAAAGTTGAGCCCAATCTT 779
213 HisThrArgLeu...HisArgGlnArgAlaLeuAspTyrGlySerGlnse 228
|||||
780 GTGACAAAGTCAACATGCCCGCGGTGCGCCAGCACCTGAACTCTCTGGG 829
228 rGlyArgGlu 231
:::
830 GGACCGTCAG 839

```

```
seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-887-352B-1
```

```
seq_documentation_block:
```

```
; Sequence 1, Application US/08887352B
```

```
; Patent No. 5994511
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardiou, John Lowe
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of
; TITLE OF INVENTION: Improving Polypeptides
```

```
; NUMBER OF SEQUENCES: 26
```

```
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Genentech, Inc.
```

```
; STREET: 1 DNA Way
```

```
; CITY: South San Francisco
```

```
; STATE: California
```

```
; COUNTRY: USA
```

```
; ZIP: 94080
```

```
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
```

```
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: WinPatin (Genentech)
```

```
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/887,352B
```

```
; FILING DATE: 03-Jul-1997
```

```
; CLASSIFICATION: 530
```

```
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Svoboda, Craig G.
```

```
; REGISTRATION NUMBER: 39,044
```

```
; REFERENCE/DOCKET NUMBER: P1123
```

```
; TELECOMMUNICATION INFORMATION:
```

```
; TELEPHONE: 650/225-1489
```

```
; TELEFAX: 650/952-9881
```

```
; INFORMATION FOR SEQ ID NO: 1:
```

```
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 6127 base pairs
```

```
; TYPE: Nucleic Acid
```

```
; STRANDEDNESS: Double
```

```
; TOPOLOGY: Circular
```

```
; US-08-887-352B-1
```

```
alignment_scores:
```

```
Quality: 99.50
```

```
Length: 259
```



```
seq_documentation_block:
; Sequence 3, Application US/07841646
; Patent No. 5266683
; GENERAL INFORMATION:
```

```
seq_documentation_block:
; Sequence 3, Application US/07841646
; Patent No. 5266683
; GENERAL INFORMATION:
```



```

; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA: US 579,865
; APPLICATION NUMBER: 07-SEP-1990
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 569,920
; FILING DATE: 20-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 483,913
; FILING DATE: 22-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 422,613
; FILING DATE: 17-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 315,342
; FILING DATE: 23-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 232,630
; FILING DATE: 15-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 179,460
; FILING DATE: 08-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3192..3730
; OTHER INFORMATION: /label= EXON-1
; OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10413..10414
; OTHER INFORMATION: /label= GAP-1
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10696..10891
; OTHER INFORMATION: /label= EXON-2
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10960..10961
; OTHER INFORMATION: /label= GAP-2
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 11059..11211
; OTHER INFORMATION: /label= EXON-3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11351..11352
; OTHER INFORMATION: /label= GAP-3
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
```

```

; NAME/KEY: exon
; LOCATION: 11420..11617
; OTHER INFORMATION: /label= EXON-4
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11721..11722
; OTHER INFORMATION: /label= GAP-4
; OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO
; OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
; OTHER INFORMATION: THIS SEQUENCE."
; FEATURE:
; NAME/KEY: exon
; LOCATION: 13354..13436
; OTHER INFORMATION: /label= EXON-5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 15044..15160
; OTHER INFORMATION: /label= EXON-6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17245..17410
; OTHER INFORMATION: /label= EXON-7
; US-08-447-570-3
```

```

alignment_scores:
  Quality: 97.00      Length: 142
  Ratio: 1.406       Gaps: 7
  Percent Similarity: 48.592   Percent Identity: 28.169

alignment_block:
US-09-135-238B-2_COPY_18_272 x US-08-447-570-3/rev ..

Align seg 1/1 to reverse of: US-08-447-570-3 from: 1 to: 17410

133 MetProAlaTyrAlaSerSerSerLysPheValThrArgValThrThrPr 149
:::||||::: :::||||::: ::: ::: :::
1975 CTTCCACCTCTCTCTCTCGGGCCCTGTATCTTTCAGACCCACCCCTCA 1926
::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
149 oAlaGlnArgGlyLysValProValHisHisSerSerProThrThrG 166
||| ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
1925 G.....CCTGAGACACACACACCCACCCACCCACCACTA 1894
166 lnleThrHisArgProArg..... 172
::: ::: ::: ::: :::
1893 GGGCTTCCCATCAGCCCTCTCTGCGACTGCTGGGAGCCAGCCGCTG 1844
173 ValSerArgAlaSerSer.ValAlaGlyAspLysProArgThrPheLeuP 189
||| ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
1843 GTCTTCATCTCTCTCTCCCTCCAGGGTCTGTGCTGAGACC..... 1801
189 roSerThrThrAlaSerLysIleSerAlaLeuGluGlyLeuLeuLysPro 205
||| ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
1800 ..TCCCATCGGGCTCTCTCTCTGTCTCTCTCATCATTCCTCTCAA 1753
206 GlnThrProSerTyrAsnHisHisThrArgLeuHisArgGln..... 219
::: ||||| ::: ::: ::: ::: ::: :::
1752 TCAATACCGAGC.....CATGCACGCTGTACAGGGAGTCCCCAGA 1712
220 .....ArgAlaLeuAspT 224
:::
1711 CCACTCAGAACCCCGGTCCTCTCTCGGTGTGGGATCTCAAGGGAAT 1662
224 yRGlySerGlnSerGlyArgGluGlyGlnGlyPhe.....HisIle 237
|| ::: ::: ::: ::: ::: ::: ::: :::
1661 ACTGACCCAGAGTGGCAGTGAGGGCAGCGGGTGGGGTAAAGTTTCATC 1612
238 LeuIleProThrIleLeuGlyLeu 245
||| ||| ::: ::: ::: ::: :::
1611 CTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
```

a

OM of: US-09-135-238B-2_COPY_18_272 to: EST:*

out_format : pfs

Date: Sep 12, 2000 6:04 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q/cgn2.1/USPTO.spool/US09135238/runat_24082000_091323_17869/app_query.fasta_1.1282
-DB=EST -QFWT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFWT=pfs -NORM=ext -MINLEN=0
-MAXLEN=100000 -USER=US09135238 -CGN1_1_2178 -NCPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-135-238B-2_COPY_18_272

Query length: 255

Database: EST:*

Database sequences: 5247842

Database length: ~2090053206

Search time (sec): 1563.530000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_est42:AW402953	+	544.00	919.78	312	! AW402953 UI-HF-BK0-abd-c-06-0-0
gb_est44:AW655210	+	451.00	756.31	518	! AW655210 105863 MARC 1BOV Bos t
gb_est42:AA444691	-	444.50	746.05	482	! AW444691 UI-H-B13-ajw-g-12-0-0U
gb_est8:AA521993	-	425.00	710.07	627	! AW521993 vntel2.r1 Soares_mamm
gb_est41:AW327067	+	415.00	699.13	415	! AW327067 20516 MARC 1BOV Bos t
gb_est45:AW655218	+	384.00	643.01	359	! AW655218 105873 MARC 1BOV Bos t
gb_est25:AI760373	-	328.00	543.94	711	! AI760373 wh87d11.x1 NCI CGAP Cl
gb_est12:AA830944	-	305.00	509.01	485	! AA830944 oc58b10.s1 NCI CGAP CG
gb_est8:AA491123	-	299.00	509.61	179	! AA491123 aa48g04.r1 NCI CGAP CG
gb_est40:AW291041	-	292.50	486.20	561	! AW291041 UI-H-B12-agj-h-06-0-0U
gb_est4:AA296671	+	281.00	469.88	417	! AA296671 vb20c07.r1 Soares_mous
gb_est3:AA174968	+	218.00	361.79	417	! AA174968 ms78h03.r1 Soares_mous
gb_est36:AW012833	+	209.50	356.43	1199	! AW012833 wh51a10.r1 Soares_mamm
gb_est15:AI021484	+	180.00	300.94	614	! AI021484 ub08h03.r1 Soares_mamm
gb_est3:AA177660	+	157.50	267.01	220	! AA177660 mt20d11.r1 Soares_mous
gb_est44:AW603666	+	145.00	245.97	728	! AW603666 CM0-CN0039-280100-168
gb_gss15:AZ033800	+	121.00	196.09	501	! AZ033800 RPCI-23-248K2.TV RPCI-
gb_est48:FI4851	+	120.50	200.58	306	! FI4851 SSC24F11 Porcine small
gb_est44:AW603303	+	120.50	193.04	614	! AW603303 RCI-CN0012-170100-011
gb_est49:HB4070	+	117.00	190.93	431	! HB4070 ys4c09.r1 Soares retina
gb_est43:AW484527	+	107.50	175.23	414	! AW484527 63153 MARC 3BOV Bos t
gb_est12:AA839051	+	105.50	170.03	489	! AA839051 vw47a02.r1 Soares_mamm
gb_est49:HB4076	+	105.00	171.06	411	! HB4076 ys94c12.r1 Soares retina
gb_est17:AI177514	+	102.50	164.20	523	! AI177514 EST221146 Normalized
gb_est39:AW176541	+	101.50	167.89	710	! C88490 C88490 Mouse early blas
gb_est40:AW258498	+	101.50	162.50	318	! AW258498 uq35h10.y1 NCI CGAP Ma
gb_est3:AA212711	+	100.50	161.82	476	! AA212711 mw81f05.r1 Soares_mous
gb_gss8:AZ034923	+	100.50	161.51	362	! AZ034923 RPCI-11-383F15.TV RPCI-
gb_gss4:AA307015	+	100.00	152.90	1003	! AA307015 HS-2046.B1 H02-T7 CIT
gb_est44:AW603280	-	99.00	157.40	566	! AW603280 RCI-CN0012-170100-011
gb_gss16:B21398	-	99.00	133.10	1066	! B21398 T8D17-Sp6 TAMU Arabidops
gb_est5:AW2194	-	96.50	145.90	2683	! AW2194 mc69a10.r1 Soares_mous
gb_est44:AW609762	-	96.00	134.50	890	! AW609762 QV0-ST0247-200100-092
gb_est46:AW110271	-	96.00	153.95	955	! AW110271 eze08ne.r1 Neurospora
gb_gss11:AO751079	+	95.50	142.19	4320	! AO751079 HS_5576.B1 G04-SP6 RF
gb_est42:AA407899	+	95.00	154.45	896	! AA407899 UI-HF-BK0-ask-f-09-0-0
gb_est15:AI022770	+	94.50	145.49	2826	! AI022770 ov99e09.x1 Soares_test
gb_gss12:AO843626	+	94.00	150.97	1401	! AO843626 LMAJFV1_lm02d09.y3 Lei
gb_est42:AW402453	+	93.50	149.16	511	! AW402453 UI-HF-BK0-aan-f-12-0-0

gb_gss4:AO324334 + 93.50 144.98 30.19 752 ! AO324334 mgxb0018E19r CUGI R
gb_est24:AI746782 + 93.00 151.96 12.33 365 ! AI746782 ul07d03.y1 Sugano m
gb_est44:AW604875 + 93.00 151.16 13.67 393 ! AW604875 QV1-CT0366-280100-0
gb_est46:AW726179 + 92.50 141.07 49.82 922 ! AW726179 GA_Ea0020006 Gossy

seq_name: gb_est42:AW402953

seq_documentation_block:
LOCUS AW402953 312 bp mRNA EST 16-FEB-2000
DEFINITION UI-HF-BK0-abd-c-06-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone
IMAGE:3055859 5', mRNA sequence.
ACCESSION AW402953
VERSION AW402953.1 GI:6921743
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 312)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138550.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M.B. Soares Lab
cDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbrrp/image/image.html
Seq primer: M13 Forward.

FEATURES
Source
1..312
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3055859"
/clone_lib="NIH_MGC_36"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/lab_host="DH10B (LT1)"
/note="Vector: pT73-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(0.5-1.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldio, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 97 a 74 c 87 g 54 t
ORIGIN

alignment_scores:
Quality: 544.00 Length: 101
Ratio: 5.386 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-135-238B-2_COPY_18_272 x AW402953
Align seg 1/1 to: AW402953 from: 1 to: 312

24 GluMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyth 40
|||||
10 GAAATGATCATGTAGGATATATCTGCCGGGAGATGGCTGGATCTGGAAC 59
40 rCysGlyThrValValSerThrThrAsnPhelLeuAlaGluTyrLysg 57
|||||
60 ATGTGTTACCGTGGTATCTCCACCACCACTTCATCAAGGAGAATAACAAGG 109
57 lYArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 73

```
|||||
110 GCGAGTCTACCTGAAACATATACCCACGCAAGATCTGTCTAGTGAG 159
74 ValThrGlnLeuThrGluSerAspSerGlyValThrAlaCysGlyAlaG1 90
|||||
160 GTAACACAGCTGACAGAAAGTGACAGCGGAGTCTATGCTGCGGAGCGG 209
90 yMetAsnThrAspArgGlyThrGlnLysValThrLeuAsnValHis 107
|||||
210 CATGAACACAGACCGGGAAAGACCCAGAAAGTCAACCTGAATGTCAC 259
107 erGlyThrGluProSerTrpGluGluGlnProMetProGluThrProLys 123
|||||
260 GTGAATACGAGCCATCATGGGAAGAGCAGCAATGCTGAGAGCTCRAA 309
124 Trp 124
|||
310 TGG 312
```

```
seq_name: gb_est45:AW655210
seq_documentation_block:
LOCUS AW655210 518 bp mRNA EST 05-APR-2000
DEFINITION 105863 MARC 1B0V Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW655210
VERSION AW655210.1 GI:7421036
KEYWORDS EST.
SOURCE Bos taurus.
ORGANISM Bos taurus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 518)
AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
Keele,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT On Mar 8, 1999 this sequence version replaced gi:4388119.
Contact: Smith RFL
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904 e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTTCCAGTCCAGCAGC
Plate: 86 row: K column: 10
Seq primer: ATTTAGTGACACTATAG.
```

```
FEATURES
source
1..518
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/Note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
```

```
BASE COUNT 124 a 133 c 157 g 104 t
ORIGIN
alignment_scores:
Quality: 451.00 Length: 119
Ratio: 4.295 Gaps: 0
Percent Similarity: 88.235 Percent Identity: 67.227
```

```
alignment_block:
US-09-135-238b-2_copy_18_272 x AW655210 ..
Align seg 1/1 to: AW655210 from: 1 to: 518
1 ArgIleLeuProGluValLysValGluGlyGluLeuGlyGlySerValTh 17
|||||
161 AAGGTCCTTCCAGAGTAAGATGGAGGAATGCTGGGAGGATCCATTAC 210
17 rIleLysCysProLeuProGluMetHisValArgIleTyrlLeuCysArg 34
|||||
211 CATCGAGTGCCTTCTCTGAACGCATGTGAGGATATATCTGTGCCGA 260
34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 50
|||||
261 CCATTGACGAGTGTGAAGATGCACACCGTGTGCCAGCAACAATAC 310
51 lleLysAlaGluTyrlLysGlyArgValThrLeuLysGlnTyPrArgly 57
|||||
311 CTCAGAGAGGAATTCAGACCCGAGTCACGCTGGAGCAGTGTCCGACAG 360
67 sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84
|||||
361 GAATCTGTCTTGTGGTGATGACAGAGCTGACCAAGAATGACAGCGGA 410
84 alTyrlAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
|||||
411 TCTATGCTGTGGGTGGCGGGAACACAGACCGAGCAAGCCAGCAGCAG 460
101 ValThrLeuAsnValHisSerGlyTyrlGluProSerTrpGluGluGlnP 117
|||||
461 ATCACCTGACTCTTCACAGCGTTTACGAGCCATCTGGGAAGAAGACC 510
117 oMetPro 119
|||||
511 GATGCCT 517
seq_name: gb_est42:AW444691
seq_documentation_block:
LOCUS AW444691 482 bp mRNA EST 17-FEB-2000
DEFINITION UI-H-B13-ajw-g-12-0-UI.sl NCI_CGAP_Sub5 Homo sapiens cDNA clone
IMAGE:2733502 3', mRNA sequence.
ACCESSION AW444691
VERSION AW444691.1 GI:6986453
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 482)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html The following repetitive
elements were found in this cDNA sequence: 1-30,
>POLY_A#Simple_repeat
Seq primer: M13 Forward
POLY_A=Yes.
Location/Qualifiers
1..482
/organism="Homo sapiens"
/db_xref="taxon:9606"
FEATURES
source
```



```

/clone_lib="IMAGE:2733502"
/clone_lib="NCI_CGAP_Sub5"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub5
is a subtracted library derived from NCI_CGAP_Sub4. The
NCI_CGAP_Sub5 library had 3 million recombinants. A
single-stranded DNA preparation of NCI_CGAP_Sub4 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)
1322376-1323911, 1456008-1456775,1500552-1502855);
NCI_CGAP_Kid5 pool 1 LLAM 3338-3342,3722-3725, 3776-3778
(IMAGE Clones) 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM,3575-3582,
3851-3854 (IMAGE Clones)
1414920-1417991,1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM
3164-3167, 3716-3720, 3733-3735 (IMAGE Clones)
1257096-1258631,1469064-1470983, 1475592-1476743);
NCI_CGAP_Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068
(IMAGE Clones) 985608-986759,1101192-1101959,
1217928-1220615); NCI_CGAP_Co10 pool 1 LLAM 2644-2653,
2871-2872 (IMAGE Clones)
1057416-1061255,1144584-1145351). (10% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clones) 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clones) 2710536-2712455) (10% of
the driver population); plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clones) 2712456-2723591) (10% of the
driver population), plus a pool of 5,472 clones from
NCI_CGAP_Sub4 (IMAGE Clones) 2723592-2728969) (70% of the
driver population). Subtraction was performed as
previously described [Bonaldo, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806. ;
TAG_LIB=NCI_CGAP_Kid5; TAG_TISSUE=kidney; TAG_SEQ=ATTC"
BASE COUNT      103 a 107 c 111 g 161 t
ORIGIN

```

```

alignment_scores:
  Quality: 444.50      Length: 117
  Ratio: 4.401        Gaps: 3
  Percent Similarity: 86.325  Percent Identity: 80.342

```

```
alignment_block:
```

```
US-09-135-238b-2_copy_18_272 x AW444691/rev ..
```

```
Align seg 1/1 to reverse of: AW444691 from: 1 to: 482
```

```

24  GluMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyTh 40
|||||
481  GNAATGATGTGAGGATATCTGTCGGGAGATGCTGTGATCTGGAAC 432
|||||
40  rCysGlyThrValValSerThrAsnPhelIleLysAlaGluTyrLysG 57
|||||
431  ATGTGGTACCGTGGTATCCACCACTTCATCAAGCCAGCAATACRAGG 382
|||||
57  lYArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 73
|||||
381  GCCGAGTTACTCTGAAGCAATACCCACCAAGAATCTGTTCTAGTGAG 332
|||||
74  ValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaG1 90
|||||
331  GTAACACAGCTGACAGAAGTGACAGCGGAGTCTATGCTCGGAGCGGG 282
|||||
90  yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 107
|||||
281  CATGAACACAGACCGGGGAAGACCCAGAAAGTCACTGATGTCCACA 232
|||||
107  er...GluTyrGluProSerTrpGluGluGlnProMetProGluThrPro 122
|||
231  GTGGTAGGTCTCGCTCGCTGATTGGAGGCTCAG.....ACTACCCAG 191

```

```

123  LysTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSe 139
|||
190  AAAATGTTCACTTGGGAAGATTG...GAACAGCCATCTAACTGTAAACAG 144
|||||

```

```

139 r 139
143 T 143

```

```
seq_name: gb_est8:AA521993
```

```

seq_documentation_block:
LOCUS      AA521993                627 bp      mRNA           EST           17-JUL-1997
DEFINITION vhl61e12.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
IMAGE:891502 5', mRNA sequence.

```

```

ACCESSION  AA521993
VERSION    AA521993.1  GI:2262738
SOURCE     EST.

```

```
ORGANISM   Mus musculus.
```

```

REFERENCE  1 (bases 1 to 627)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,K. and
Waterston,K.

```

```

TITLE      The WashU-HHMI Mouse EST Project
JOURNAL    Unpublished (1996)
COMMENT    On Jan 14, 1998 this sequence version replaced gi:1797443.

```

```

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

```

```

Email: mouseest@watson.wustl.edu

```

```

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:519462
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 494.

```

```
FEATURES
```

```
source
```

```

1..627
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:891502"
/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Organ: mammary gland; Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCGAATGGTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

```

```

BASE COUNT      165 a 192 c 156 g 114 t
ORIGIN

```

```
alignment_scores:
```

```

  Quality: 425.00      Length: 184
  Ratio: 3.148        Gaps: 5
  Percent Similarity: 73.370  Percent Identity: 51.087

```

```
alignment_block:
US-09-135-238B-2_COPY_18_272 x AA521993  ..
Align seg 1/1 to: AA521993 from: 1 to: 627

87 CysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLe 103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10 TGTGTGTGGGCATGAGACACACAAGGCAAGACCCAGAAATCACCT 59
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 uAsnValHisSerGluTyr...GluProSerTrpGluGluGlnProMetP 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 GAATTCATTAAGTAATACCCAGAACCAATTCCTGGAGATGAATGGACCT 109
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119 roGluThrProLysTrpPheHisLeuProLysLeuPheGlnMetPro... 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 CTGAGCGGCCAAGATGGTGCACAGATTTCTGCAGCACCAGATGCCCTGG 159
135 .....AlaTyrAlaSerSerSerLysPheValThrArgValTh 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 CTCCACGGGAGTGAACATCCAGCTCTTCTGGAGTCATAGCAAGATTAC 209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
147 rThrProAlaGlnArgGlyLysValProValHisHisSerSerProt 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 CAGCGCACCTCAAGACTGAGGCCCTCCGGTTCACGAGCCCTCCAGCA 259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 hrThrGlnIleThrHisArgProArgValSerArgAlaSerSerVala 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
260 TCACCTTCAGTAACCAACATCCAGAGTTTACAGAGCATTTTCTGTGTCA 309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 GlyAspLysProArgThrPheLeuProSerThrThrAlaSerLysIleSe 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 GCTACCAAGTCCCGCGCTCTGCCAGCAACACAGCTCAAGACTTC 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
197 rAlaLeuGlyGlyLeuLysProGlnThrPro.....SerTyrAsnH 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 CACTCAGCAGGCAAT.....CAGGCCCTAGAGCGCAGCTACAGCC 399
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
212 iHisThrArgLeuHisArgGlnAlaLeuAspTyrGlySerGlnSer 228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 ACCACACAGACTTCATGAGCAAGGACACGCCACCATGGCCCACTAT 449
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
229 GlyArgGluGlyGlnGly.....PheHisIleLe 238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
450 GGGAGAGAAGCCGAGGGCTTCACATCCCATCCCAAGAAATTTACATCT 499
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 uileProThrIleLeuGlyLysPheLeuAlaLeuLysValVal 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
500 GATTCGACCTTCTCGGCTTCTCTTCTGTTCTTTTGGGACTGTGG 549
255 al 255
||
550 TA 551

seq_name: gb_est41:AW327067

seq_documentation_block:
LOCUS AW327067 359 bp mRNA EST 27-JAN-2000
DEFINITION 20516 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW327067
VERSION AW327067.1 GI:6762988
KEYWORDS EST.
SOURCE Bos taurus.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 359)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
Keele,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
```

```
JOURNAL Unpublished (2000)
COMMENT On Jan 6, 2000 this sequence version replaced gi:6676592.
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 13 row: K column: 5
Seq primer: ATTAGGTGACACTATAG.
FEATURES
Location/Qualifiers
1..359
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 90 a 111 c 87 g 71 t
ORIGIN
alignment_scores:
Quality: 415.00 Length: 119
Ratio: 4.192 Gaps: 1
Percent Similarity: 83.193 Percent Identity: 64.706
alignment_block:
US-09-135-238B-2_COPY_18_272 x AW327067 ..
Align seg 1/1 to: AW327067 from: 1 to: 359
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2 GTCACGCTGGAGCAGTGTCCGGACAGGAATCTGCTTGTGGTGATGAC 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
52 AGAGCTGACCAAGAATGACAGCGGATCTATGCTGTGGGTGGCGGGA 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102 ACACAGACCGGCAAGACCCAGCAGATCACCTGACTGTTTCACAGCGTT 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 TyrGluProSerTrpGluGluGlnProMetProGluThrProLysTrpPh 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 TACGAGCCATCTCTGGGAAGAAGACCGATGCTGAGCCTCCAGCATGGTT 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 e.....HisLeuProTyrLeuPheGlnMetProAlaT 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 TAATAGATTTTACAAATGCACTTGCTCTCTGGTTCAGATGCTGACAC 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 yAlaSerSerSerLysPheValThrArgValThrProAlaGlnArg 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252 ATGCCAGTTCTTTAAATTCACGTCCAAAAGTAACCAACACCAAGG 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 GlyLysValProValHisHisSerSerProThrThrGlnIleThrH 169
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 ATCAAGTCTCTCGAGCACACCGGCTCCCCCAACCCCTCAGTCACCCA 351
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 sArgPro 171
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
352 CCGCCCC 358
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: gb_est45:AW65218
```

```

seq_documentation_block:
LOCUS      AW655218             494 bp      mRNA           EST           05-APR-2000
DEFINITION 1058733 MARC IBOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION  AW655218
VERSION    AW655218.1   GI:7421044
KEYWORDS   EST.
SOURCE     Bos taurus.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
REFERENCE  Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
AUTHORS    Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
            Keeler,J.W.
TITLE      Design and use of four pooled tissue normalized cDNA libraries for
JOURNAL    EST discovery in cattle
COMMENT    EST discovered (2000)
            Unpublished (2000)
            On Mar 8, 1999 this sequence version replaced gi:4388128.
            Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAACAGCATGACCAT
            BACKWARD: GTTTCCTCCAGTCACGACG
            Plate: 86 row: M column: 8
            Seq primer: ATTAGGTGACACTATAG.
FEATURES   Location/Qualifiers
            source          1..494
                        /organism="Bos taurus"
                        /db_xref="taxon:9913"
                        /clone_lib="MARC IBOV"
                        /tissue_type="pooled"
                        /lab_host="DH10B"
                        /note="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                        Library made from pooled tissue from lymph node, ovary,
                        fat, hypothalamus, and pituitary."
BASE COUNT      118 a      125 c      150 g      101 t
ORIGIN
alignment_scores:
            Quality: 384.00      Length: 111
            Ratio: 4.042
            Percent Similarity: 85.586      Percent Identity: 63.063
alignment_block:
US-09-135-238B-2_COPY_18_272 x AW655218      ..
Align seg 1/1 to: AW655218 from: 1 to: 494
1 ArgIleLeuProGluValLysGluGlyLeuGlyGlySerValTh 17
162 AGGTCTCTCCAGAGTAAGATGGAAGGAAATGCTGGGAGGATCCATTAC 211
17 rIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArgG 34
212 CATCGAGTCCCACTTCCTGAAACGCAATGAGGATATATCTGTGCCGGA 261
34 LuMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
262 CCATTGACGAGTCTGGAGATGACCAACCGTGTCTCCAGCAACAATAC 311
51 IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67
312 GTAAGAGAGATTCAAGCACCAGCGCTGGAGCAGCTGTCGGGACAG 361

```

```

67 SASLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84
362 GAATCTGTCTCTGTGGTGACAGAGCTGACCAAGATGACAGCGGA 411
84 alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
412 TCTATGCTGTGGGTGGCGGGAACACAGACCGAGCGGACACCATCAG 461
101 ValThrLeuAsnValHisSerGluTyrGluPro 111
462 ATCACCTGACTGTTACAGCGTTTACTAGCCA 494
seq_name: gb_est25:AI760373
seq_documentation_block:
LOCUS      AI760373             711 bp      mRNA           EST           20-DEC-1999
DEFINITION wh87d11.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2387733 3'
            similar to TR:O60667 O60667 ANTI-FAS-INDUCED APOPTOSIS. ; mRNA
            sequence.
ACCESSION  AI760373
VERSION    AI760373.1   GI:5176040
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 711)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    On Oct 30, 1998 this sequence version replaced gi:3812483.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
            CDNA Library Preparation: M. Bento Soares, Ph.D.
            CDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1449 Std Error: 0.00
            Seq primer: -400P from Gibco
            High quality sequence stop: 472.
FEATURES   Location/Qualifiers
            source          1..711
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone_image="IMAGE:2387733"
                        /clone_lib="NCI_CGAP_CLL1"
                        /tissue_type="B-cell, chronic lymphocytic leukemia"
                        /lab_host="DH10B"
                        /note="vector: pT7T3D-Pac (Pharmacia) with a modified
                        polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                        was primed with a Not I - oligo(dT) primer [5',
                        TGTACCAATCTGAAGTGGAGCGCGCATGCTGTTTTTTTTTTTTTTTTTTT
                        T 3']; double-stranded cDNA was ligated to Eco RI
                        adaptors (Pharmacia), digested with Not I and cloned into
                        the Not I and Eco RI sites of the modified pT7T3 vector.
                        Library is normalized, and was constructed by Bento
                        Soares and M.Fatima Bonaldo."
BASE COUNT      127 a      195 c      239 g      148 t
ORIGIN
alignment_scores:
            Quality: 328.00      Length: 84
            Ratio: 4.316
            Percent Similarity: 90.476      Percent Identity: 83.333

```

```
alignment_block:
US-09-135-238B-2_COPY_18_272 x A1760373/rev ..
Align seg 1/1 to reverse of: A1760373 from: 1 to: 711

173 ValSerArgAlaSerSerValAlaGlyAspLysProArgThrPheLeuPr 189
711 GTGTCAGAGCAATCTCAAGTAGCAGGTGACAAAGCCGACCTCTTTGCC 662

189 oSerThrThraLysSerAlaLeuGluGlyLeuLeuLysProG 206
|||||
561 ATCCACTACAGCTCAAAATCTCAGCTCTGGAGGGGCTGCTCAAGCCCC 612
|||||
206 lnThrPro.SerTyrAsnHisHisThrArgLeuHisArgGlnAlaLe 222
|||||
611 AGACGCCAGGGGTACACACACCACCAGCGCTGCACAGCGAGAGCACT 562

222 uAspTyrGlySerGlnSerGlyArgGluGlyGlnGlyPheHisIleLeuI 239
|||||
561 GGACTATGGCTCAGAGTCTGGAGGAGGAGCCAGAGATTTCACATCCTGA 512

239 leProThrIleLeuGlyLeuPheLeuAlaLeuLeuGlyLeuValVal 255
|||||
511 TCCCGACCATCTGGGGCTTTTCTGCTGGCACTTCTGGGGTGGTGGTG 462

seq_name: gb_est12:AA830944

seq_documentation_block:
LOCUS AA830944 485 bp mRNA EST 07-APR-1998
DEFINITION OC58B10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1353883 3',
mRNA sequence.
ACCESSION AA830944
VERSION AA830944.1 GI:2904043
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 485)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 983 Std Error: 0.00
Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 473.
Location/Qualifiers
1. 485
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1353883"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer"

[5'-TGTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

```
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD+),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTCACATCTCAAGTGGAGCGCCCTCATTTTTTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

BASE COUNT 46 a 69 c 34 g 30 t

ORIGIN

alignment_scores:

Quality:	299.00	Length:	59
Ratio:	5.068	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-09-135-238B-2_COPY_18_272 x AA491123 ..

Align seg 1/1 to: AA491123 from: 1 to: 179

146 ValThrThrProAlaGlnArgGlyLysValProValHisHisSerSe 162
|||||
3 GTTACACACACAGCTCAAAGGGCAAGTCCCTCCAGTTCACCATCTC 52
162 rProThrThrGlnIleThrHisArgProArgValSerArgAlaSerV 179
|||||
53 CCCCACACCCCAAAATCACCCACCGACCTCGAGTGTCACAGCATCTTCAG 102
179 alAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLys 195
103 TAGCAGGTGCACAGCCCGAACCTTCCTGGCCATCCACTACAGCTCAAAA 152
196 IleSerAlaLeuGluGlyLeuLeuLys 204
|||||
153 ATCTCAGCTCTGGAGGGGTGCTCAAG 179

seq_name: gb_est40:AW291041

seq_documentation_block:

LOCUS AW291041 561 bp mRNA EST 16-JAN-2000
DEFINITION UT-H-B12-agt1-h-06-0-UI.sl NCI_CGAP_Sub4 Homo sapiens cDNA clone
IMAGE:2724707 3', mRNA sequence.
ACCESSION AW291041
VERSION AW291041.1 GI:6697677
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 561)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On May 18, 1998 this sequence version replaced gi:3137211.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares lab Clone distribution:
NCI-CGAP clone distribution information can be found through the

I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source

Location/Qualifiers

1..561

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2724707"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NCI_CGAP_Sub4 library is a subtracted library derived from
the NCI_CGAP_Sub2 library which is a subtracted library
derived from the NCI_CGAP_Sub1 library, which is a
subtracted library derived from B1. B1 constitutes a
mixture of 21 normalized or subtracted NCI_CGAP
libraries: NCI_CGAP_Co4, NCI_CGAP_Pr22, NCI_CGAP_Pr28,
NCI_CGAP_Co10, NCI_CGAP_Co16, NCI_CGAP_Kid5,
NCI_CGAP_Kid12, NCI_CGAP_Kid3, NCI_CGAP_Kid11,
NCI_CGAP_Lym2, NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_CLL1,
NCI_CGAP_Le12, NCI_CGAP_Brn23, NCI_CGAP_Lu5,
NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
NCI_CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI_CGAP_Kid3 pool 1 : LLAM 3334-3337, 3682-3683,
3798-3803 (IMAGE CloneIDs 1322376-1323911,
1456008-1456775, 1500552-1502855) NCI_CGAP_Kid5 pool 1 :
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE CloneIDs
132912-1329831, 1471368-1472903, 1492104-1493255)
NCI_CGAP_Lu5 pool 1 : LLAM 3575-3582, 3851-3854 (IMAGE
CloneIDs 1414920-1417991, 1520904-1522439) NCI_CGAP_GC4
pool 1 : LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
CloneIDs 1257096-1258631, 1468064-1470983,
1475592-1476743) NCI_CGAP_Pr22 pool 1 : LLAM 2457-2459,
2758-2759, 3062-3068 (IMAGE CloneIDs 985608-986759,
1101192-1101959, 1217928-1220615) NCI_CGAP_Co10 pool 1 :
LLAM 2644-2653, 2871-2872 (IMAGE CloneIDs 1057416-1061255,
1144584-1145351) Subtraction was performed as previously
described [Bonaldo, Lennon & Soares (1996): Normalization
and Subtraction: Two Approaches To Facilitate Gene
Discovery. Genome Research 6, 791-806.];
TAG_L1B=NCI_CGAP_Co10; TAG_TISSUE=colon; TAG_SEQ=AAACG"

BASE COUNT 121 a 125 c 120 g 194 t 1 others

ORIGIN

alignment_scores:

Quality:	292.50	Length:	91
Ratio:	3.953	Gaps:	3
Percent Similarity:	81.319	Percent Identity:	72.527

alignment_block:

US-09-135-238B-2_COPY_18_272 x AW291041/rev ..

Align seg 1/1 to reverse of: AW291041 from: 1 to: 561

50 PheIleLysAlaGluThrLysGlyArgValThrLeuLysGluThrProAr 66
|||||
561 TTCATCAGGCAGATACAGCCCGAGTTACTCTGAAGCAATACCCAGC 512
66 glysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83
|||||
511 CAAGAATCTGTTCTTAGTGGAGTAACTACAGCTGACAGAAAGTGACACG 462
83 lyValThrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
|||||
461 GAGTCTATGCTCGCGAGCGGCATGAACACACAGACCGGNGAAGACCCAG 412
100 LysValThrLeuAsnValHisSer...GluThrGluProSerTrpGluCl 115

|||||
411 AAAGTCAACCTGAATGTCACAGTGGTAGGTCTCTCCGCTGATTGGAGGC 362
115 uGlnProMetProGluThrProLysTrpPheHisLeuProTyrlLeuPheG 132
361 TCAG.....ACTACCCAGAAAATGTTTCACCTGGGAAGATTG...G 324
132 lnMetProAlaTyrlAlaSerSer 139
323 AACAGCCATCTAACTGTAACAGT 301

alignment_block:
US-09-135-238B-2_COPY_18_272 x AA289667 ..
Align seg 1/1 to: AA289667 from: 1 to: 417
145 ArgValThrProAlaGlnArgGlyLysValProProValHisHis 161
4 AAAGTTACACGCCCAACTCCAAAGACTGAGGCCCTCGGTTCCACGCC 53
161 rSerProThrThrGlnIleThrHisArgProArgValSerArgAlaSerS 178
54 CTCACGATCATCTTCAGTAAACCAATCCAGAGTTTACAGAGCATTTT 103
178 erValAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSer 194
104 CTGTGTCAAGTACCAGTCCCAAGCGCTCTCCAGCAACACAGCCTCA 153
195 LysIleSerAlaLeuGluGlyLeuLysProGlnThrProSerTyrlAs 211
154 AAGACTTCCACTCAGCAAGCA...ATCAGGCCCTTAGAGCCAGCTACAG 200
211 nHisHisThrArgLeuHisArgGlnArgAlaLeuAspTyrlGlySerGlnS 228
201 CCACCAACACGACGCTTCATGAGCAAGAGACGCCACCTGGGCCACACT 250
228 erGlyArgGluGlyGlnGly.....PheHisIle 237
251 ATGGGAGAGAAGACCGGGCTTCACATCCCATCCAGAAATTCACATC 300
238 LeulleProThrIleLeuGlyLeuPheLeuAlaLeuGlyLeuVa 254
301 CTGATCCGACCTCTCGGGCTTCTCTCTGCTGGTCTTTTGGGACTGGT 350
254 lVal 255
351 GGTA 354

seq_name: gb_est4:AA289667
seq_documentation_block:
LOCUS AA289667 417 bp mRNA EST 14-APR-1997
DEFINITION v020c07.r1 Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:749484
5', mRNA sequence.
ACCESSION AA289667
VERSION AA289667.1 GI:1937343
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 417)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,F., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:458468
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 405.
Location/Qualifiers
1. .417
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:749484"
/clone_lib="Soares mouse 3NDMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dt) primer [5',
TGTTACCAATCTGAAGTGGGAGCGCGCTGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."

BASE COUNT 109 a 140 c 91 g 77 t
ORIGIN
alignment_scores:
Quality: 281.00 Length: 118
Ratio: 3.122 Gaps: 2
Percent Similarity: 76.271 Percent Identity: 50.847

seq_name: gb_est3:AA174968
seq_documentation_block:
LOCUS AA174968 460 bp mRNA EST 16-FEB-1997
DEFINITION ms78h03.r1 Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:617717
5', mRNA sequence.
ACCESSION AA174968
VERSION AA174968.1 GI:1756098
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 460)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,F., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:378541
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 459.
Location/Qualifiers
1. .460

FEATURES
source
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:749484"
/clone_lib="Soares mouse 3NDMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dt) primer [5',
TGTTACCAATCTGAAGTGGGAGCGCGCTGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."

seq_name: gb_est3:AA174968
seq_documentation_block:
LOCUS AA174968 460 bp mRNA EST 16-FEB-1997
DEFINITION ms78h03.r1 Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:617717
5', mRNA sequence.
ACCESSION AA174968
VERSION AA174968.1 GI:1756098
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 460)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,F., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:378541
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 459.
Location/Qualifiers
1. .460

FEATURES
source
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:749484"
/clone_lib="Soares mouse 3NDMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dt) primer [5',
TGTTACCAATCTGAAGTGGGAGCGCGCTGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Fatima Bonaldo."


```

199 A 199
seq_name: gb_est36:AW012833

seq_documentation_block:
LOCUS AW012833      327 bp    mRNA          EST              10-SEP-1999
DEFINITION uc03dl1.v1 Sugano mouse liver mlia Mus musculus cDNA clone
IMAGE:2582325 5' similar to gb:U06431 Mus musculus B6/CBAFJ
polymorphic immunoglobulin receptor mRNA, (MOUSE);,, mRNA sequence.
ACCESSION AW012833
VERSION   AW012833.1 GI:5861611
KEYWORDS house mouse.
SOURCE    house mouse.
ORGANISM  Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 327)
AUTHORS  Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Ritter,E., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Waterston,R. and Willson,R.
TITLE     The WashU-NCI Mouse EST Project 1999
JOURNAL   Unpublished (1999)
COMMENT   On Jul 9, 1999 This sequence version replaced gi:5433821.
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:1021913
Seq primer: custom primer used
High quality sequence stop: 272.
Location/Qualifiers
FEATURES             source
source               1..327
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone_image=IMAGE:2582325"
                     /clone_lib="Sugano mouse liver mlia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
(CACCATGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGGGCCCTTTTTTTTTTTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTCGCCTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACATGTG, 3' site CACATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTCCTCGCTCAAGAGTGC and 3' end
primer GCACCTGCACCTCGAGACA."
BASE COUNT        87 a 98 c 76 g 66 t
ORIGIN
alignment_scores:
Quality: 180.00 Length: 88
Ratio: 2.769 Gaps: 3
Percent Similarity: 73.864 Percent Identity: 40.909

alignment_block:
US-09-135-238B-2_COPY18_272 x AW012833 ..
Align seg 1/1 to: AW012833 from: 1 to: 327

```


TGTTACCAATCTGAAGTGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTT
T 3'; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldi."

BASE COUNT 62 a 78 c 42 g 38 t

ORIGIN

alignment_scores:

Quality: 157.50 Length: 68
Ratio: 3.088 Gaps: 1
Percent Similarity: 75.000 Percent Identity: 50.000

alignment_block:

US-09-135-238B-2_COPY_18_272 x AI021484 ..

Align seg 1/1 to: AI021484 from: 1 to: 220

173 ValSerArgAlaSerSerValAlaGlyAspLysProArgThrPheLeuPr 189

1 GTTTACAGACATTCTGTGTCAGCTACCAAGTCCCGAGCGCTCTGCC 50

189 oSerThrAlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProG 206

51 AGCAACACACAGCCTCAAGACTTCCACTCAGCAAGCA...ATCAGGCCCC 97

206 lnThrProSerTyrAsnHisHisThrArgLeuHisArglnArgAlaLeu 222

98 TAGAGGCCAGCTACAGCCACACACAGACTTTCATGAGCAAGGACACGC 147

223 AspTyrGlySerGlnSerGlyArgGluGlyGlnGlyPheHisIleLeuI 239

148 CACCATGGCCCCACACTATGGGAGAGAGACCGAGGCGCTTCACATCCCCAT 197

239 ePro 240

198 CCA 201

OM of: US-09-135-238B-2_COPY_273_390 to: GenEmbl:* out_format : pfs
Date: Sep 12, 2000 6:56 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame-p2n.model -DEV=xlp
-O=/cgn2_1/uspto_spol/US09135238/runat_24082000_091323_17880/app_query.fasta_1.1282
-DB=GenEmbl -QEMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=1000000
-USER=US09135238@cgn1_1_5120 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-135-238B-2_COPY_273_390
Query length: 118
Database: GenEmbl:*
Database sequences: 972840
Database length: 892348106
Search time (sec): 2784.300000

score_list:	Strd	Orig	Zscore	EScore	Len	Documentation
gb_pr3:AF057557	+	616.00	637.82	2.0e+27	1339	AF057557 Homo sapiens anti-Fas
gb_htg11:AC023534	+	325.00	308.02	4.6e+09	221365	AC023534 Homo sapiens clone
gb_htg20:AC010538	-	107.00	85.28	1.2e+04	249227	AC010538 Homo sapiens chromo
gb_ba1:SC135	-	105.00	95.10	3.4e+03	490909	AL031541 Streptomyces coelic
gb_htg14:AC025634	+	104.50	91.25	5.5e+03	68092	AC025634 Homo sapiens chromo
gb_pr4:AC020663	+	104.50	85.89	1.1e+04	154218	AC020663 Homo sapiens chromo
gb_htg22:AC027687	-	104.50	82.83	1.6e+04	245472	AC027687 Homo sapiens chromo
gb_ba1:SCF43A	-	104.00	95.03	3.4e+03	35437	AL096837 Streptomyces coelic
gb_ba1:SVPKS	-	103.50	106.38	790.84	5820	X16300 Streptomyces violaceoru
gb_pr1:AB019004	-	103.50	105.42	893.49	6728	AB019004 Homo sapiens DNA, 5'
gb_ba1:SVGRA	-	103.50	105.36	900.75	6793	X16144 S. violaceoruber DNA fo
gb_ro:RTPRMTNP2	+	103.50	101.01	1.6e+03	13187	Z46939 R. norvegicus PRM1, PRM
gb_ba1:SCB63	+	103.50	94.20	3.8e+03	37200	AL035640 Streptomyces coelic
gb_ba1:SVT011500	-	103.50	93.85	3.9e+03	32550	AJ011500 Streptomyces violace
gb_pr3:HS16369	-	103.50	85.35	1.2e+04	143291	AL008733 Human DNA sequence
gb_pr1:HUMFGRB	-	103.00	114.54	277.63	1553	L78720 Human fibroblast growth
gb_htg8:AC020867	+	103.00	80.59	2.1e+03	273785	AC020867 Mus musculus clone
gb_ba2:AE001874	-	102.50	101.19	1.5e+03	10991	AF001874 Deinococcus radiodur
gb_pr3:HS74782	-	102.00	85.57	1.1e+04	109859	AL021393 Human DNA sequence
gb_ba2:AF071081	+	101.50	106.89	740.36	3946	AF071081 Mycobacterium tubercu
gb_ba1:AC93410	+	101.50	106.89	740.36	3946	A93410 Sequence 6 from Patent
gb_ba1:SCNUSG	-	101.50	102.91	1.2e+03	7235	X72787 S. griseus nusG, rplKJL
gb_pr4:AC004973	-	101.50	85.87	1.1e+04	97037	AC004973 Homo sapiens PAC clc
gb_pr3:HUMYWD703	-	101.50	83.70	1.4e+04	135038	L78810 Homo sapiens ADP/ATP
gb_ro:AF082568	+	101.00	103.67	1.1e+03	5968	AF082568 Mus musculus zinc fin
gb_htg5:AC016278	+	101.00	91.90	5.1e+03	35816	AC016278 Homo sapiens clone R
gb_ba1:SC5C7	+	101.00	90.87	5.8e+03	41906	AL031515 Streptomyces coelic
gb_ba1:SCM1	-	101.00	90.56	6.0e+03	43961	AL133422 Streptomyces coelic
gb_htg27:AC022536	+	101.00	81.89	1.8e+04	164587	AC022536 Homo sapiens clone
gb_htg18:AC025300	-	101.00	80.74	1.9e+04	168628	AC025300 Homo sapiens chromo
gb_pr4:AC004169	+	101.00	80.96	2.0e+04	189805	AC004169 Homo sapiens chromo
gb_htg20:AC024549	+	101.00	79.96	2.3e+04	220961	AC024549 Homo sapiens chromo
gb_pr4:AF152486	+	100.50	109.01	563.85	2445	AF152486 Homo sapiens protocad
gb_pr4:AF152316	+	100.50	103.99	1.1e+03	5260	AF152316 Homo sapiens protocad
gb_ba1:SC6D7	-	100.50	91.05	5.6e+03	37750	AL133213 Streptomyces coelic
gb_htg18:AC026593	-	100.50	89.41	7.0e+03	48459	AC026593 Homo sapiens chromo
gb_htg20:AC008468	+	100.50	83.47	1.5e+04	119783	AC008468 Homo sapiens chromo
gb_pr3:AC005609	-	100.50	81.65	1.9e+04	157970	AC005609 Homo sapiens chromo
gb_htg2:AL157817	-	100.50	80.65	2.1e+04	184106	AL157817 Homo sapiens chromo

gb_pr1:HUMHRARC + 100.00 110.15 487.42 1903 ! M24857 Human retinoic acid
gb_ba1:STMSEC - 100.00 109.91 502.89 1975 ! D17464 S. griseus genes for
gb_pr1:AB018254 - 100.00 101.88 1.4e+03 6706 ! AB018254 Homo sapiens mRNA
gb_ba1:SC9A10 - 100.00 99.95 1.8e+03 9000 ! AL031260 Streptomyces coeli
gb_ba1:MTCX10D7 - 100.00 90.19 6.3e+03 39800 ! Z79700 Mycobacterium tuber
seq_name: gb_pr3:AF057557
seq_documentation_block:
LOCUS AF057557 1339 bp mRNA PRI 04-JUN-1998
DEFINITION Homo sapiens anti-Fas-induced apoptosis (TOSO) mRNA, complete cds.
ACCESSION AF057557
VERSION AF057557.1 GI:3169292
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1339)
AUTHORS Hitoshi,Y., Lorens,J., Kitada,S.I., Fisher,J., LaBarge,M.,
Ring,H.Z., Francke,U., Reed,J.C., Kinoshita,S. and Nolan,G.P.
TITLE Toso, a cell surface, specific regulator of Fas-induced apoptosis
in T cells
JOURNAL Immunity 8 (4), 461-471 (1998)
MEDLINE 98246048
REFERENCE
2 (bases 1 to 1339)
AUTHORS LaBarge,M. and Hitoshi,Y.
TITLE Direct Submission
JOURNAL Submitted (03-APR-1998) Molecular Pharmacology, Stanford
University, 300 Pasteur Dr., Palo Alto, CA 94304, USA
FEATURES
Location/Qualifiers
1..1339
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="lq31-q32"
/cell_line="Jurkat"
/cell_type="T cell"
/note="TOSO.2 cDNA; cloned by PCR using primers designed
from the original TOSO.1 cDNA sequence"
1..1339
/gene="TOSO"
/note="transmembrane protein; Ig superfamily member"
/codon_start=1
/product="anti-Fas-induced apoptosis"
/protein_id="AAC18830.1"
/db_xref="GI:3169293"
variation
127
/gene="TOSO"
/note="occurs in TOSO.1 cDNA"
/replace="a"
1015
/gene="TOSO"
/note="occurs in TOSO.1 cDNA"
/replace="g" 340 g 258 t
BASE COUNT 309 a 432 c 340 g 258 t
ORIGIN
alignment_scores:
Quality: 616.00 Length: 118
Ratio: 5.220 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

```
US-09-135-238B-2_COPY_273_390 x AF057557      ..
Align seg 1/1 to: AF057557 from: 1 to: 1339

1  LysArgAlaValGluArgGlyAlaLeuSerArgAlaArgArgLe 17
   |||||
836  AAAAGGCCGCTGAAAGGAGGAAAGCCCTCTCCAGCGGGCCCGGACT 885
   |||||

17  uAlaValArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerP 34
   |||||
886  GCGCGTGAGATCGCGCCCTGGAGAGCTCCAGAGGCCCGCGGGTCGC 935
   |||||

34  rGATGProArgSerGlnAsnAsnIleTyrSerAlaCysProArgArgAla 50
   |||||
936  CGCGAGCCGCTCCCAAAACAACATCTACAGCGCTGCCGCGCGCGCT 985
   |||||

51  ArgGlyAlaAspAlaAlaGlyThrGlyGluAlaProValProGlyProG 67
   |||||
986  CGTGAGCGGACCTCGAGGCACAGGGAAGCCCGCTTCCCGGCCCGCG 1035
   |||||

67  yAlaProLeuProAlaProLeuGlnValSerGluSerProTrrpLeuH 84
   |||||
1036  AGCGCGGTGCGCCCGCGCGCTGCAGGTGCTGAATCTCCTGGCTCC 1085
   |||||

84  IsAlaProSerLeuLysThrSerCysGluTyrValSerLeuTyrHisGln 100
   |||||
1086  ATGCCCATCTCTGAAGACCACTGTGAATCTGAGCTCTACACCAG 1135
   |||||

101  ProAlaAlaMetMetGluAspSerAspSerAspTyrIleAsnValPr 117
   |||||
1136  CCTGCGCCCATGTAGGAGCAGTTCATTCAGATGACTACATCAATGTTC 1185
   |||||

117  oAla 118
   ||||
1186  TGCC 1189
```

seq_name: gb_btgl1:AC023534

```
seq_documentation_block:
LOCUS AC023534 221365 bp DNA HTG 02-MAR-2000
DEFINITION Homo sapiens clone RP11-462N18, *** SEQUENCING IN PROGRESS ***, 61
unordered pieces.
ACCESSION AC023534
VERSION AC023534.2 GI:7143452
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 221365)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome, clone RP11-462N18
Unpublished
2 (bases 1 to 221365)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeArelano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,Y., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisan,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
```

Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2000 this sequence version replaced gi:6978210.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5244
Center clone name: 462_N18

* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1108: contig of 1108 bp in length
* 109 1208: gap of 100 bp
* 1209 2292: contig of 1084 bp in length
* 2293 2392: gap of 100 bp
* 2393 3673: contig of 1281 bp in length
* 3674 3773: gap of 100 bp
* 3774 4947: contig of 1174 bp in length
* 4948 5047: gap of 100 bp
* 5048 6104: contig of 1057 bp in length
* 6105 6204: gap of 100 bp
* 6205 7259: contig of 1035 bp in length
* 7260 7359: gap of 100 bp
* 7360 8667: contig of 1308 bp in length
* 8668 8767: gap of 100 bp
* 8768 10566: contig of 1799 bp in length
* 10567 10666: gap of 100 bp
* 10667 12173: contig of 1507 bp in length
* 12174 12273: gap of 100 bp
* 12274 13643: contig of 1370 bp in length
* 13644 13743: gap of 100 bp
* 13744 15215: contig of 1472 bp in length
* 15216 15315: gap of 100 bp
* 15316 16609: contig of 1294 bp in length
* 16610 16709: gap of 100 bp
* 16710 18085: contig of 1376 bp in length
* 18086 18185: gap of 100 bp
* 18186 19668: contig of 1483 bp in length
* 19669 19768: gap of 100 bp
* 19769 22083: contig of 2315 bp in length
* 22084 22183: gap of 100 bp
* 22184 24052: contig of 1869 bp in length
* 24053 24152: gap of 100 bp
* 24153 25902: contig of 1750 bp in length
* 25903 26002: gap of 100 bp
* 26003 27639: contig of 1637 bp in length
* 27640 27739: gap of 100 bp
* 27740 29377: contig of 1638 bp in length
* 29378 29477: gap of 100 bp
* 29478 30910: contig of 1433 bp in length
* 30911 31010: gap of 100 bp
* 31011 31225: contig of 215 bp in length
* 31226 31325: gap of 100 bp
* 31326 32890: contig of 1565 bp in length
* 32891 32990: gap of 100 bp
* 32991 34812: contig of 1822 bp in length
* 34813 34912: gap of 100 bp

```
* 34913 36817: contig of 1905 bp in length
* 36818 36917: gap of 100 bp
* 36918 38575: contig of 1658 bp in length
* 38576 38675: gap of 100 bp
* 38676 41309: contig of 2634 bp in length
* 41310 41409: gap of 100 bp
* 41410 43344: contig of 1935 bp in length
* 43345 43444: gap of 100 bp
* 43445 45953: contig of 2409 bp in length
* 45954 45953: gap of 100 bp
* 45954 48434: contig of 2481 bp in length
* 48435 48534: gap of 100 bp
* 48535 51026: contig of 2492 bp in length
* 51027 51126: gap of 100 bp
* 51127 54369: contig of 3243 bp in length
* 54370 54469: gap of 100 bp
* 54470 56661: contig of 2192 bp in length
* 56662 56761: gap of 100 bp
* 56762 59097: contig of 2336 bp in length
* 59098 59197: gap of 100 bp
* 59198 61551: contig of 2354 bp in length
* 61552 61651: gap of 100 bp
* 61652 64257: contig of 2606 bp in length
* 64258 64357: gap of 100 bp
* 64358 67511: contig of 3154 bp in length
* 67512 67611: gap of 100 bp
* 67612 70329: contig of 2718 bp in length
* 70330 70429: gap of 100 bp
* 70430 73572: contig of 3143 bp in length
* 73573 73672: gap of 100 bp
* 73673 77203: contig of 3531 bp in length
* 77204 77303: gap of 100 bp
* 77304 80599: contig of 3296 bp in length
* 80600 80699: gap of 100 bp
* 80700 84046: contig of 3347 bp in length
* 84047 84146: gap of 100 bp
* 84147 88357: contig of 4211 bp in length
* 88358 88457: gap of 100 bp
* 88458 92454: contig of 3997 bp in length
* 92455 92554: gap of 100 bp
* 92555 96892: contig of 4338 bp in length
* 96893 96992: gap of 100 bp
* 96993 101334: contig of 4342 bp in length
* 101335 101434: gap of 100 bp
* 101435 105438: contig of 4004 bp in length
* 105439 105538: gap of 100 bp
* 105539 110413: contig of 4875 bp in length
* 110414 110513: gap of 100 bp
* 110514 115994: contig of 5481 bp in length
* 115995 116094: gap of 100 bp
* 116095 120447: contig of 4353 bp in length
* 120448 120547: gap of 100 bp
* 120548 124483: contig of 3936 bp in length
* 124484 124583: gap of 100 bp
* 124584 130249: contig of 5666 bp in length
* 130250 130349: gap of 100 bp
* 130350 135462: contig of 5113 bp in length
* 135463 135562: gap of 100 bp
* 135563 139847: contig of 4285 bp in length
* 139848 139947: gap of 100 bp
* 139948 146731: contig of 6784 bp in length
* 146732 146831: gap of 100 bp
* 146832 151469: contig of 4638 bp in length
* 151470 151569: gap of 100 bp
* 151570 159901: contig of 8332 bp in length
* 159902 160001: gap of 100 bp
* 160002 168308: contig of 8307 bp in length
* 168309 168408: gap of 100 bp
* 168409 176195: contig of 7787 bp in length
* 176196 176295: gap of 100 bp
* 176296 189303: contig of 13008 bp in length
* 189304 189403: gap of 100 bp
* 189404 202135: contig of 12732 bp in length
```

```
* 202136 202235: gap of 100 bp
* 202236 221365: contig of 19130 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..221365
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="RP11-462N18"
                /clone_lib="RPCI-11 Human Male BAC"
            1..1108
                /note="assembly_fragment"
            1209..2292
                /note="assembly_fragment"
            2393..3673
                /note="assembly_fragment"
            3774..4947
                /note="assembly_fragment"
            5048..6104
                /note="assembly_fragment"
            6205..7259
                /note="assembly_fragment"
            7360..8667
                /note="assembly_fragment"

alignment_scores:
    Quality: 325.00      Length: 144
    Ratio: 4.333        Gaps: 1
    Percent Similarity: 52.083      Percent Identity: 50.000

alignment_block:
    US-09-135-238B-2_COPY_273_390 x AC023534 ..
    Align seg 1/1 to: AC023534 from: 1 to: 221365

      8 LysAlaLeuSerArgAlaArgAlaValArgLeuAlaValArgMetArgAlaLeu 24
      :::::::::::::::::::::::::::::::::::::::::::::::::::::
    32391 GAAGCCCTCTCCAGGCGGGCGCCGCGGCTGGCGGTGAGGATCGCGCCCT 32440
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::

      24 uGluSerSerGlnArgProArgGlySerProArgProArgSerGlnAsnA 41
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
    32441 GGAGAGCTCCAGAGAGGCCCGCGGGGTGCGCGGACCGCGCTCCCAACACA 32490
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::

      41 snlleTyrSerAlaCysProArgArgAlaArgGlyAlaAspAlaAla... 56
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
    32491 ACATCTACAGCGCCTCCCGCGGCGCGCTCGTGTGAGCGGAGCGCTCGAGGT 32540
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::

      56 ..... 56
    32541 GAGCCCGCGGGCGGCCACCGGGGAGTAGCGGAAAGAGTGCCTGGCTCCG 32590
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::

      56 ..... 56
    32591 GGAAGCGCGGGCGGACGAGCGGGCGCGCTCCAGGTGCCAGGGCTTCCACACA 32640
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::

      56 ..... 56
    32641 GGTGTGAGCTGGGAGTGGGGCGGGGTGGAGCTGGGAGTAGTAGCATGAC 32690
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::

      56 ..... 56
    32691 GGTGGGGCGGGAGCGCTCGGCGCGCGCCTCACCTCGCCCTGTCGTCCT 32740
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::

      57 GlyThrGlyGluAlaProValProGlyProGlyAlaProLeuProProA 73
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
    32741 AGCACAGGGGAGGCGCCCGTCCCGGCGCGGAGCGCGGTGGCCCCCG 32790
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::

      73 laProLeuGlnValSerGluSerProTrpLeu 83
      ::::::::::::::::::::::::::::::::::::::::::::::
    32791 CCCCGCTGCAGGTAAAGCTCCGCTCTCGCCCTG 32822
      ::::::::::::::::::::::::::::::::::::::::::::::

seq_name: gb.htg20:AC010538
seq_documentation_block:
    LOCUS AC010538 249227 bp DNA HTG
    DEFINITION Homo sapiens chromosome 16 clone RP11-36817, WORKING DRAFT
```

28-MAR-2000

SEQUENCE, 90 unordered pieces.
AC010538
VERSION AC010538.3 GI:7431338
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 249227)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
REFERENCE 2 Unpublished
2 (bases 1 to 249227)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Mar 28, 2000 this sequence version replaced gi:6758794.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>
-----Summary Statistics
Consensus quality: 189807 bases at least Q40
Consensus quality: 22221 bases at least Q30
Consensus quality: 230312 bases at least Q20
Estimated insert size: 245950; agarose-fp estimation
Estimated insert size: 245950; agarose-fp estimation
Quality coverage: 3.16x in Q20 bases; agarose-fp estimation
Quality coverage: 3.12x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 90 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1060: contig of 1060 bp in length
* gap of unknown length
* 1061 2237: contig of 1177 bp in length
* gap of unknown length
* 2238 3270: contig of 1033 bp in length
* gap of unknown length
* 3271 4550: contig of 1280 bp in length
* gap of unknown length
* 4551 5647: contig of 1097 bp in length
* gap of unknown length
* 5648 6804: contig of 1157 bp in length
* gap of unknown length
* 6805 7807: contig of 1003 bp in length
* gap of unknown length
* 7808 9086: contig of 1279 bp in length
* gap of unknown length
* 9087 10101: contig of 1015 bp in length
* gap of unknown length
* 10102 11251: contig of 1150 bp in length
* gap of unknown length
* 11252 12446: contig of 1195 bp in length
* gap of unknown length
* 12447 13557: contig of 1111 bp in length
* gap of unknown length
* 13558 14576: contig of 1019 bp in length
* gap of unknown length
* 14577 15705: contig of 1129 bp in length
* gap of unknown length
* 15706 17020: contig of 1315 bp in length
* gap of unknown length
* 17021 18385: contig of 1365 bp in length
* gap of unknown length
* 18386 19680: contig of 1295 bp in length
* gap of unknown length
* 19681 20882: contig of 1202 bp in length
* gap of unknown length
* 20883 21954: contig of 1072 bp in length
* gap of unknown length
* 21955 23689: contig of 1735 bp in length
* gap of unknown length
* 23690 24962: contig of 1273 bp in length
* gap of unknown length
* 24963 26319: contig of 1357 bp in length
* gap of unknown length
* 26320 27582: contig of 1263 bp in length
* gap of unknown length
* 27583 28849: contig of 1267 bp in length
* gap of unknown length
* 28850 29861: contig of 1012 bp in length
* gap of unknown length
* 29862 31003: contig of 1142 bp in length
* gap of unknown length
* 31004 32367: contig of 1364 bp in length
* gap of unknown length
* 32368 33785: contig of 1418 bp in length
* gap of unknown length
* 33786 34834: contig of 1049 bp in length
* gap of unknown length
* 34835 35941: contig of 1107 bp in length
* gap of unknown length
* 35942 37397: contig of 1456 bp in length
* gap of unknown length
* 37398 38403: contig of 1006 bp in length
* gap of unknown length
* 38404 39501: contig of 1098 bp in length
* gap of unknown length
* 39502 40724: contig of 1223 bp in length
* gap of unknown length
* 40725 41991: contig of 1267 bp in length
* gap of unknown length
* 41992 43426: contig of 1435 bp in length
* gap of unknown length
* 43427 45236: contig of 1810 bp in length
* gap of unknown length
* 45237 46594: contig of 1458 bp in length
* gap of unknown length
* 46595 47746: contig of 1052 bp in length
* gap of unknown length
* 47747 49191: contig of 1445 bp in length
* gap of unknown length
* 49192 51044: contig of 1853 bp in length
* gap of unknown length
* 51045 52378: contig of 1334 bp in length
* gap of unknown length
* 52379 53505: contig of 1127 bp in length
* gap of unknown length
* 53506 55286: contig of 1781 bp in length
* gap of unknown length
* 55287 56793: contig of 1507 bp in length
* gap of unknown length
* 56794 58628: contig of 1835 bp in length
* gap of unknown length
* 58629 59765: contig of 1137 bp in length
* gap of unknown length
* 59766 61158: contig of 1393 bp in length
* gap of unknown length
* 61159 62688: contig of 1530 bp in length
* gap of unknown length
* 62689 64788: contig of 2100 bp in length
* gap of unknown length
* 64789 66536: contig of 1748 bp in length
* gap of unknown length
* 66537 70463: contig of 3927 bp in length
* gap of unknown length
* 70464 72740: contig of 2277 bp in length
* gap of unknown length
* 72741 75083: contig of 2343 bp in length
* gap of unknown length

```

* 75084 gap of unknown length
* 77658: contig of 2575 bp in length
* 77659 gap of unknown length
* 79607: contig of 1949 bp in length
* 79608 gap of unknown length
* 81867: contig of 2260 bp in length
* 81868 gap of unknown length
* 84131: contig of 2264 bp in length
* 84132 gap of unknown length
* 86256: contig of 2125 bp in length
* 86257 gap of unknown length
* 88488: contig of 2232 bp in length
* 88489 gap of unknown length
* 90327: contig of 1839 bp in length
* 90328 gap of unknown length
* 92118: contig of 1791 bp in length
* 92119 gap of unknown length
* 94665: contig of 2547 bp in length
* 94666 gap of unknown length
* 97095: contig of 2430 bp in length
* 97096 gap of unknown length
* 99715: contig of 2620 bp in length
* 99716 gap of unknown length
* 102136: contig of 2421 bp in length
* 102137 gap of unknown length
* 104714: contig of 2578 bp in length
* 104715 gap of unknown length
* 107130: contig of 2416 bp in length
* 107131 gap of unknown length
* 109573: contig of 2443 bp in length
* 109574 gap of unknown length
* 112859: contig of 3286 bp in length
* 112860 gap of unknown length
* 115291: contig of 2432 bp in length
* 115292 gap of unknown length
* 117865: contig of 2574 bp in length
* 117866 gap of unknown length
* 121526: contig of 3661 bp in length
* 121527 gap of unknown length
* 125987: contig of 4461 bp in length
* 125988 gap of unknown length
* 132407: contig of 8420 bp in length
* 132408 gap of unknown length
* 137033: contig of 4626 bp in length
* 137034 gap of unknown length
* 141135: contig of 4102 bp in length
* 141136 gap of unknown length
* 146850: contig of 5715 bp in length
* 146851 gap of unknown length
* 152336: contig of 5486 bp in length
* 152337 gap of unknown length
* 159908: contig of 7572 bp in length

alignment_scores:
  Quality: 107.00      Length: 84
  Ratio: 2.378        Gaps: 4
  Percent similarity: 53.571  Percent identity: 38.095

alignment_block:
US-09-135-238b-2_COPY_273_390 x AC010538/rev ..

Align seg 1/1 to reverse of: AC010538 from: 1 to: 249227

32 GlySerProArgSerGlnAsnIleTyrSerAlaCysProAr 48
|||||:|||||:|||||:|||||:|||||:|||||:|||||
112713 GGAACCCCGGCGCCGACACCCAGACCGGCTACCCAGCCCG 112664
48 q.....:.....:.....:.....:.....:.....:.....
      |.....:.....:.....:.....:.....:.....:.....
112663 GCACCCCGGCGGCGACCCCGACAGAGGGCTCGACCGCCGAGGCAC 112614
55 .....AlaAlaGlyThrGlyGlu.....:.....:.....:.....:.....
      ||| ||||| ||| :|||:|||||:|||||:|||||:|||||

```

```

112613 CCGCAGCCCGGAGAACCCCGACCGGTGGCTCACCAGCCCGCCAGCTCCC 112564
67 GlyAlaProLeuProAlaProLeuGlnValSerGluSerProTrpLe 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||
112563 CCAGCCCGGCTCCCGACCGCGCTCTCCAGCCCGCCAGTCCAGTGGCT 112514
83 uHisAlaProSerLeuLysThrSerCysGluTyrValSerLeuTyrHisG 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||
112513 AGGTATCCCTAAG.....CACCTGGCTCAGGTGAGCACTCACCATC 112473
100 ln 100
||
112472 AA 112471

seq_name: gb_bal:SCI35

seq_documentation_block:
LOCUS SCI35 40909 bp DNA BCT 09-SEP-1998
DEFINITION Streptomyces coelicolor cosmid I35.
ACCESSION AL031541
VERSION AL031541.1 GI:3581833
KEYWORDS 3-hydroxyacyl-coa dehydrogenase; 50S ribosomal protein L20; 50S
ribosomal protein L35; arabinosidase; argC; argJ; ATC initiation
codon; ATP/GTP binding protein; dehydrogenase; glutamate
N-acetyltransferase; glutamine synthetase; infC; IS117; mutT
domain; N-acetyl-gamma-glutamyl-phosphate reductase;
phenylalanyl-tRNA synthetase; pheS; pheT; rplI; rpmI; rRNA
methylase; secreted serine protease; transcriptional regulator;
translation initiation factor IF-3; transposase; TTA leu codon;
two-component sensor.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2).
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 40909)
AUTHORS Oliver, K. and Harris, D.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 40909)
AUTHORS Parkhill, J., Barrell, B. G. and Rajandream, M. A.
JOURNAL Direct Submission
TITLE Submitted (08-SEP-1998) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
REFERENCE 3 (bases 1 to 40909)
AUTHORS Redenbach, M., Kieser, H. M., Denapaite, D., Eichner, A., Cullum, J.,
Kinashi, H. and Hopwood, D. A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
COMMENT
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S\_coelicolor/) CDS are
numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons using a
specially developed Hidden Markov Model (Krogh et al., Nucleic
Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
of Bibb et al., Gene 30:157-66(1984) as implemented at
http://www.nih.gov.jp/

```

jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 135 lies between 110 and 141 on the AseI-I genomic restriction fragment.

FEATURES

source

```
1. 40909
  /organism="Streptomyces coelicolor A3(2)"
  /strain="A3(2)"
  /db_xref="taxon:100226"
  /clone="cosmid 135"
  complement(1..552)
  /gene="argJ"
  complement(<1..552)
  /note="argJ"
```

gene

CDS

```
/note="SCI35.01c, argJ, probable glutamate
N-acetyltransferase, partial CDS, len: >184 aa: similar to
many e.g. ARGJ CORGL glutamate N-acetyltransferase
(ornithine acetyltransferase) (EC 2.3.1.35) (388 aa),
fasta scores; opt: 614 z-score: 909.0 E(): 0, 55.4%
identity in 184 aa overlap. Referred to as 'argB' on
genetic map"
/codon_start=1
/transl_table=11
/product="putative glutamate N-acetyltransferase"
/protein_id="CAA20790.1"
/db_xref="GI:3581834"
/db_xref="SPTREMBL:O88040"
```

gene

CDS

```
/translation="MSVTAAGFTAGITAGIKESGSPDIALVVTGPPRRSAGVFTIS
NRVKAAPLWSEQLKSGEVTAVVNLGSGNACTGPKGFQDTHATAEKADVLGTGAG
EYAVCSGLIGVLLPMDKLPGVEAAGQLSEHGGEKAAIAKTTDTHKTSVVTROG
WTVGMAKGAGMLAPGLATMLVI"
complement(549..1577)
/gene="argC"
complement(549..1577)
/gene="argC"
/note="SCI35.02c, argC, N-acetyl-gamma-glutamyl-phosphate
reductase, len: 342 aa; identical to ARGCSWRCO
N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38)
(249 aa). Contains PS01224
N-acetyl-gamma-glutamyl-phosphate reductase active site.
Referred to as 'argA' on genetic map."
/codon_start=1
/transl_table=11
/product="N-acetyl-gamma-glutamyl-phosphate reductase"
/protein_id="CAA20791.1"
/db_xref="GI:3581835"
```

RBS

```
/translation="MAVRAAVAGSAGYAGGELLRLTLHPVEIGALTGNSNAGORLG
ALQPHLLADRLVLEATTPEVLGGHDVVFLALPHGGSAAVEQJGPDVLYVWDGADR
LKADGWERYFSGPHAGTPYGLPELPGAALAEKRIAPVGCYPTAVSLALFAYIA
ASLAPEPAVIAASGTSAGAKAPKPHLLGSEVNGSPYGGGGRHTEPMIONLGAV
ACEPVTSFTPLAPMPRGILATCTAKAKPGVTAESVRAAYEKALADEPFFHLLPEGO
WPATASVYGSNAVQVQVADHAAAGRIITAISAINDLAKGTAGGAVQSNLALGLDTEG
LTITIGAP"
complement(560..563)
/gene="argC"
/note="possible RBS upstream of argJ"
complement(831..1877)
/note="99.8% match to SCARGC X66783 S.coelicolor argC gene
(partial)"
complement(1107..1157)
```

misc_feature

```
/gene="argC"
/note="PS01224 N-acetyl-gamma-glutamyl-phosphate reductase
active site"
complement(1585..1590)
/note="possible RBS upstream of argC"
1611..2174
```

CDS

```
/gene="SCI35.03"
1611..2174
/note="SCI35.03"
/note="SCI35.03, unknown, len: 187 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCI35.03"
/protein_id="CAA20792.1"
/db_xref="GI:3581836"
/db_xref="SPTREMBL:O88041"
/transl_table="MPMHVYAMRCGVTSRRRRRLACRFRTPSTVGVAIVAVIVELY
VGLRWQRAALVYSLCALALPLWFAVNTGTRYGQGTAVTEGFDCEWDGCPGAWRL
ADGSLGRGIDGDTAFEEELVEDIPLFGGRDVAWADRGTLRLHAVAQYTVAPVGAAL
VYRIARRSAYNQETTALMRYRTTP"
complement(2164..2688)
/gene="SCI35.04c"
complement(2164..2688)
/gene="SCI35.04c"
/note="SCI35.04c, ATP/GTP binding protein, len: 174 aa;
similar to TMRB_BACSU tunicamycin resistance protein (197
aa), fasta scores; opt: 282 z-score: 488.2 E(): 6.1e-20,
39.5% identity in 119 aa overlap. Contains PS00017
ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="ATP/GTP binding protein"
/protein_id="CAA20793.1"
/db_xref="GI:3581837"
/db_xref="SPTREMBL:O88042"
/transl_table="MIVWLNGHAGKTTGALVQRLLPDSRVDAEKVGTLMDDITP
GLPFDVNFQHWAPWRPLVETARVLDITGGTLVMPMTVLVEEYWRFAAGLAAHGVP
VRHFVNHADQETLRGTRAGDTVLGSPDFRLRYLPEYAEARTWLHAEAEVVDTHLT
PAEALRTAEAVRA"
complement(2647..2670)
/gene="SCI35.04c"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(2751..5225)
/gene="SCI35.05c"
complement(2751..5225)
/gene="SCI35.05c"
/note="SCI35.05c, probable secreted arabinosidase, len:
824 aa: similar to e.g. TR:O59218 (EMBL:U15178)
arabinosidase (EC 3.2.1.55) (Bacteroides ovatus) (660 aa),
fasta scores; opt: 1204 z-score: 1030.7 E(): 0, 41.4%
identity in 601 aa overlap"
/codon_start=1
/transl_table=11
/product="putative secreted arabinosidase"
/protein_id="CAA20794.1"
/db_xref="GI:3581838"
/db_xref="SPTREMBL:O88043"
```

misc_feature

```
/translation="MSKTRWRGLSVTLTAALLPSPHAEVADVITIVDPSDRG
PAIDDTMGVFEYEDINRAADGGLYAEVLVONRSFEYTDNASTYPLTAAWADGAEVY
NDGRLNERNRYLSLAAGSTVTNAGYNTGIRVEKERYDFSVWRAGHGTLTLTTLT
DAAGTLAKAROVAAGKRWKYTATFTATRSRGLAVTTTAPALDMVSLFPRDTYR
HQPGKIKDLAEKTEALHPGLFPGGLVNTGSMEDYSADSGWKRSYQWKDTIGP
VEERANANFWGYNQSGLGTYEYFEAEEDIGAMLPVVPALVTCGONKAVDDALL
KRHIQTDLLIEFANGPTSEWGKRAEMGHKPFHLTHLEVGNEENTPKFEARFEQ
FRYIAEKYDPDVTYVNSGPDGATFTDFAQLARDAGVDWDEHYEYVNSPFWLQNDY
RYSYDRNGPKVFLGYSQGNKALAEAFMTNNTGDRVFPSTATGTPSYSGVGLST
VOWSPDIWNNHASWSANTYEQKLFNNTGDRVFPSTATGTPSYSGVGLST
WATSAAYDDVKVTSADGETLLSDDFSGDASRWASGAGSWTVQDGOYVQTDAAAEHTM
VTAGDPADHYDLHVRAATKKSGEGLVAFVQKDTGNYWNNLGGWNTOSAVQASD
GGKSTLKVAGSIEFGAYDIDVYKRGQVTLYLDDGEWGSGFTDDKPAEPQVTRD
DRGTELLIVWNAQDARTADVLGGKAYASRAAVTILADQADVNTGIDAPVPTVTS
TFSGAASEFTYFPANSVTELRQR"
complement(5231..5234)
/note="possible RBS upstream of SCI35.05c"
5337..5340
/note="possible RBS upstream of SCI35.06"
5362..5820
```

alignment_scores:


```
* 19954 20773: contig of 820 bp in length
* 20774 20873: gap of 100 bp
* 20874 21666: contig of 793 bp in length
* 21667 21766: gap of 100 bp
* 21767 22560: contig of 794 bp in length
* 22561 22660: gap of 100 bp
* 22661 23494: contig of 834 bp in length
* 23495 23594: gap of 100 bp
* 23595 24390: contig of 796 bp in length
* 24391 24490: gap of 100 bp
* 24491 25300: contig of 810 bp in length
* 25301 25400: gap of 100 bp
* 25401 26245: contig of 845 bp in length
* 26246 26345: gap of 100 bp
* 26346 27146: contig of 801 bp in length
* 27147 27246: gap of 100 bp
* 27247 28051: contig of 805 bp in length
* 28052 28151: gap of 100 bp
* 28152 28974: contig of 823 bp in length
* 28975 29074: gap of 100 bp
* 29075 29894: contig of 820 bp in length
* 29895 29994: gap of 100 bp
* 29995 30822: contig of 828 bp in length
* 30823 30922: gap of 100 bp
* 30923 31753: contig of 831 bp in length
* 31754 31853: gap of 100 bp
* 31854 32644: contig of 791 bp in length
* 32645 32744: gap of 100 bp
* 32745 33564: contig of 820 bp in length
* 33565 33664: gap of 100 bp
* 33665 34453: contig of 789 bp in length
* 34454 34553: gap of 100 bp
* 34554 35334: contig of 781 bp in length
* 35335 35434: gap of 100 bp
* 35435 36239: contig of 805 bp in length
* 36240 36339: gap of 100 bp
* 36340 37167: contig of 828 bp in length
* 37168 37267: gap of 100 bp
* 37268 38063: contig of 796 bp in length
* 38064 38163: gap of 100 bp
* 38164 38947: contig of 784 bp in length
* 38948 39047: gap of 100 bp
* 39048 39835: contig of 788 bp in length
* 39836 39935: gap of 100 bp
* 39936 40761: contig of 826 bp in length
* 40762 40861: gap of 100 bp
* 40862 41671: contig of 810 bp in length
* 41672 41771: gap of 100 bp
* 41772 42598: contig of 827 bp in length
* 42599 42698: gap of 100 bp
* 42699 43500: contig of 802 bp in length
* 43501 43600: gap of 100 bp
* 43601 44405: contig of 805 bp in length
* 44406 44505: gap of 100 bp
* 44506 45324: contig of 819 bp in length
* 45325 45424: gap of 100 bp
* 45425 46219: contig of 795 bp in length
* 46220 46319: gap of 100 bp
* 46320 47120: contig of 801 bp in length
* 47121 47220: gap of 100 bp
* 47221 48051: contig of 831 bp in length
* 48052 48151: gap of 100 bp
* 48152 48967: contig of 816 bp in length
* 48968 49067: gap of 100 bp
* 49068 49877: contig of 810 bp in length
* 49878 49977: gap of 100 bp
* 49978 50789: contig of 812 bp in length
* 50790 50889: gap of 100 bp
* 50890 51669: contig of 780 bp in length
* 51670 51769: gap of 100 bp
* 51770 52573: contig of 804 bp in length
* 52574 52673: gap of 100 bp
* 52674 53475: contig of 802 bp in length
```

```
* 53476 53575: gap of 100 bp
* 53576 54421: contig of 846 bp in length
* 54422 54521: gap of 100 bp
* 54522 55331: contig of 810 bp in length
* 55332 55431: gap of 100 bp
* 55432 56226: contig of 795 bp in length
* 56227 56326: gap of 100 bp
* 56327 57139: contig of 813 bp in length
* 57140 57239: gap of 100 bp
* 57240 58075: contig of 836 bp in length
* 58076 58175: gap of 100 bp
* 58176 58991: contig of 816 bp in length
* 58992 59091: gap of 100 bp
* 59092 59911: contig of 820 bp in length
* 59912 60011: gap of 100 bp
* 60012 60819: contig of 808 bp in length
* 60820 60919: gap of 100 bp
* 60920 61699: contig of 780 bp in length
* 61700 61799: gap of 100 bp
* 61800 62605: contig of 806 bp in length
* 62606 62705: gap of 100 bp

alignment_scores:
    Quality: 104.50      Length: 81
    Ratio: 2.090        Gaps: 5
    Percent Similarity: 61.728      Percent Identity: 39.506

alignment_block:
US-09-135-238B-2_COPY_273_390 x AC025634  ..
Align seg 1/1 to: AC025634 from: 1 to: 68092

26 SerSerGlnArgProArgGlySerProArgPro.....ArgSerGlnAs 40
|||||
38457 TCCTCGAGCGCCCGAGAGATGCCCGCCCGCGCTCGCTCAGAC 38506
40 nAsnIleTyrSerAlaCysProArgArgAlaArgGly.....AlaAspA 55
|||
38507 GCGGTGCTCGCCCGCCCGCCCAAGAGCTCGCGGGGGGTGTGCCCGC 38556
|||||
55 laAlaGlyThrGlyGluAlaProValProGlyProGlyAlaPro..... 69
|||||
38557 CGCGCGCCCGGGAACGCTCCGCGTCCGTGAGGACCCCGCGTGT 38606
70 ...LeuProAlaProLeuGlnValSerGluSerProTyrLeuHisAl 85
|||||
38607 GGGCTCCCGCGCGCGCTCTCAGATGTCGCGGGCCGCC.....GC 38647
85 aProSerLeuLysThrSerCysGluTyrValSerLeuTyrHis 99
|||||
38648 CCGGNC...CGGCGAGGTGTGCGGGGCCACAGTCGACAT 38687

seq_name: gb_pr4.AC020663

seq_documentation_block:
LOCUS AC020663 154218 bp DNA PRI 08-JAN-2000
DEFINITION Homo sapiens chromosome 16 clone RPCI-11_127120, complete sequence.
ACCESSION AC020663
VERSION AC020663.1 GI:6682593
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 154218)
AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
Campbell,J.C., Fawcett,J., Maitbie,M., Bussod,M., Sutherland,R.,
Han,C. and Deaven,L.
TITLE Sequencing of Human Chromosome 16q12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 154218)
```

AUTHORS Ricke,D.O.
 TITLE Large Scale Sequence Analysis and Annotation with the Sequence
 JOURNAL Comparison Analysis (SCAN) System
 REFERENCE Unpublished
 AUTHORS 3 (bases 1 to 154218)
 Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
 Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
 Bryant,J., Tesmer,J., Melnick,L., Longmire,J., White,S., Tatum,O.,
 Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
 McMurry,K., Han,C. and Deaven,L.
 TITLE Direct Submission
 JOURNAL Submitted (08-JAN-2000) Center for Human Genome Studies, DOE Joint
 Genome Institute, Los Alamos National Laboratory, MS M888, Los
 Alamos, NM 87545, USA
 FEATURES Location/Qualifiers
 source
 1. .154218
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RPC1-11.127120"
 complement(41. .330)
 /rpt_family="Alu"
 repeat_region 575. .882
 repeat_region 1414. .1712
 repeat_region 1595. .1862
 repeat_region 1871. .2190
 repeat_region 2176. .2372
 repeat_region 2498. .2795
 repeat_region 3187. .3504
 repeat_region complement(3571. .3890)
 repeat_region /rpt_family="Alu"
 repeat_region complement(3859. .4163)
 repeat_region 4198. .4503
 repeat_region /rpt_family="Alu"
 repeat_region 4826. .5155
 repeat_region /rpt_family="Alu"
 repeat_region 5749. .6045
 repeat_region /rpt_family="Alu"
 misc_feature complement(6848. .6983)
 repeat_region 7473. .7776
 repeat_region /rpt_family="Alu"
 repeat_region 7931. .8385
 repeat_region /rpt_family="Alu"
 repeat_region 8808. .9139
 repeat_region /rpt_family="Alu"
 repeat_region 9394. .9682
 repeat_region /rpt_family="Alu"
 repeat_region 9515. .9812
 repeat_region /rpt_family="Alu"
 repeat_region 9702. .9987
 repeat_region /rpt_family="Alu"
 repeat_region 9965. .10332
 repeat_region /rpt_family="Alu"
 repeat_region complement(10403. .10835)
 repeat_region /rpt_family="Alu"
 repeat_region complement(10801. .10977)
 repeat_region /rpt_family="Alu"
 repeat_region 11580. .11780
 repeat_region /rpt_family="MER42"
 repeat_region 11803. .12101
 repeat_region /rpt_family="Alu"
 repeat_region 12282. .12576
 repeat_region /rpt_family="Alu"
 repeat_region 12643. .13073

repeat_region /rpt_family="Alu"
 12923. .13237
 repeat_region /rpt_family="Alu"
 13420. .13725
 repeat_region /rpt_family="Alu"
 13761. .14148
 repeat_region /rpt_family="Alu"
 14072. .14371
 repeat_region /rpt_family="Alu"
 complement(14401. .14696)
 repeat_region /rpt_family="Alu"
 15061. .15366
 repeat_region /rpt_family="Alu"
 15383. .15706
 repeat_region /rpt_family="Alu"
 complement(16083. .16185)
 repeat_region /rpt_family="L1"
 complement(16192. .16347)
 repeat_region /rpt_family="Alu"
 16359. .16831
 repeat_region /rpt_family="Alu"
 complement(18420. .18718)
 repeat_region /rpt_family="Alu"
 complement(18518. .18917)
 repeat_region /rpt_family="Alu"
 complement(18754. .19036)
 repeat_region /rpt_family="Alu"
 complement(19456. .19559)
 repeat_region /note="GRAIL 2 excellent exon, frame 1"
 20239. .20397
 repeat_region /rpt_family="MER20"
 complement(20808. .21104)
 repeat_region /rpt_family="Alu"
 complement(22116. .22443)
 repeat_region /rpt_family="Alu"
 complement(22789. .22948)
 repeat_region /note="GRAIL 2 excellent exon, frame 1"
 23346. .23646
 repeat_region /rpt_family="Alu"
 complement(24046. .24341)
 repeat_region /rpt_family="Alu"
 25205. .25461
 repeat_region /rpt_family="Alu"
 25605. .25862
 repeat_region /rpt_family="L1"
 25914. .25981
 repeat_region /rpt_family="Alu"
 complement(26044. .26355)
 repeat_region /rpt_family="Alu"
 28437. .28732
 repeat_region /rpt_family="Alu"
 30597. .30772
 misc_feature /note="GRAIL 2 excellent exon, frame 2"
 31300. .31655
 repeat_region /rpt_family="Alu"
 32059. .32124
 repeat_region /rpt_family="Alu"
 complement(32302. .32492)
 repeat_region /rpt_family="Alu"
 32507. .32839
 repeat_region /rpt_family="Alu"
 33649. .34040
 repeat_region /rpt_family="Alu"
 complement(34202. .34278)
 repeat_region /rpt_family="L1"
 34279. .34556
 repeat_region /rpt_family="Alu"
 35017. .35495
 repeat_region /rpt_family="Alu"
 complement(36639. .36991)
 repeat_region /rpt_family="Alu"
 37667. .37808
 misc_feature /note="GRAIL 2 excellent exon, frame 0"

```
repeat_region complement(37920..37961)
repeat_region /rpt_family="MIR"
repeat_region 38430..38745
repeat_region /rpt_family="Alu"
repeat_region 38789..38899
repeat_region /rpt_family="MER3"
repeat_region complement(40939..41089)
repeat_region /rpt_family="Alu"
repeat_region 42294..42589
repeat_region /rpt_family="Alu"
repeat_region complement(43093..43396)
repeat_region /rpt_family="Alu"
repeat_region complement(43274..43695)
repeat_region /rpt_family="Alu"
repeat_region 44405..44482
repeat_region /rpt_family="MLTIC"
repeat_region 44515..44693
repeat_region /rpt_family="Alu"
repeat_region 45816..46254
repeat_region /rpt_family="Alu"
misc_feature complement(46300..46347)
repeat_region /note="GRAU 2 excellent exon, frame 1"
repeat_region 46426..46596
repeat_region /rpt_family="MER20"
repeat_region 47139..47448
repeat_region /rpt_family="Alu"
repeat_region 47535..47943
repeat_region /rpt_family="HE1"
repeat_region complement(47939..48377)
repeat_region /rpt_family="Alu"
repeat_region 48431..48778
repeat_region /rpt_family="L1"
repeat_region 48877..49260
repeat_region /rpt_family="Alu"
repeat_region complement(49310..49588)
repeat_region /rpt_family="Alu"

alignment_scores:
  Quality: 104.50      Length: 96
  Ratio: 2.133        Gaps: 4
  Percent Similarity: 51.042  Percent Identity: 36.458

alignment_block:
US-09-135-238B-2_COPY_273_390 x AC020663 ..
Align seg 1/1 to: AC020663 from: 1 to: 154218

      2 ArgAlaValGluArgArg..... 7
140217 CGCGCGCGGAGCGGACGACTCCCGCTCTCCCGCTCCCTTCGGCTC 140266
      8 .....LysAlaLeuSerArgArgAlaArgAlaVal 19
140267 GTGACACGAGCGCGCGGCTGTGAGCGCGCGCGCGCGCGGTGCG 140316
      19 alArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerProArg 35
140317 ACCGCTGTAGCGGAGAGGAGC.....CGGCTCGCGCTCGCGCCGC 140360
      36 ProArgSerGlnAsnAsnIleTyrSerAlaCysPro.....ArgArg 49
140361 CCCCCG.....GTCTTGGACTCCGCGCCCTCCCTCTCGGCG 140398
      49 gAlaArgGlyAlaAspAlaGlyThrGlyGluAlaProValProGlyP 66
140399 CTTCCGTTTACCGCGGTGTGTCGCGCGCGCGCGCGCGGCGCATTC 140448
      66 roGlyAlaProLeuProProAlaProLeuInValSer 78
140449 CCGAGCGCTGTCGCGCGCTCGCAGCGCGCGAGTGAGC 140486
seq_name: gb_htg22:AC027687
```

```
seq_documentation_block:
LOCUS AC027687 245472 bp DNA HTG 01-APR-2000
DEFINITION Homo sapiens chromosome 16 clone RP11-6114, WORKING DRAFT SEQUENCE,
50 unordered pieces.
ACCESSION AC027687
VERSION AC027687.1 GI:7381751
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 245472)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 245472)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (01-APR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Consensus quality: 177301 bases at least Q40
Consensus quality: 204545 bases at least Q30
Consensus quality: 220813 bases at least Q20
Estimated insert size: 245472; sum-of-contigs estimation
Estimated insert size: 169000; agarose-fp estimation
Quality coverage: 7.37x in Q20 bases; agarose-fp estimation
Quality coverage: 5.08x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 50 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1298: contig of 1298 bp in length
* gap of unknown length
* 1299 2303: contig of 1005 bp in length
* gap of unknown length
* 2304 3390: contig of 1087 bp in length
* gap of unknown length
* 3391 4442: contig of 1052 bp in length
* gap of unknown length
* 4443 5615: contig of 1173 bp in length
* gap of unknown length
* 5616 6822: contig of 1207 bp in length
* gap of unknown length
* 6823 8057: contig of 1235 bp in length
* gap of unknown length
* 8058 9328: contig of 1271 bp in length
* gap of unknown length
* 9329 11175: contig of 1847 bp in length
* gap of unknown length
* 11176 12414: contig of 1239 bp in length
* gap of unknown length
* 12415 13648: contig of 1234 bp in length
* gap of unknown length
* 13649 14938: contig of 1290 bp in length
* gap of unknown length
* 14939 16002: contig of 1064 bp in length
* gap of unknown length
* 16003 17497: contig of 1495 bp in length
* gap of unknown length
* 17498 19150: contig of 1653 bp in length
* gap of unknown length
* 19151 20166: contig of 1016 bp in length
* gap of unknown length
```

```

* 20167 21607: contig of 1441 bp in length
*      gap of unknown length
* 21608 22911: contig of 1304 bp in length
*      gap of unknown length
* 22912 24331: contig of 1420 bp in length
*      gap of unknown length
* 24332 26162: contig of 1831 bp in length
*      gap of unknown length
* 26163 28064: contig of 1902 bp in length
*      gap of unknown length
* 28065 29929: contig of 1865 bp in length
*      gap of unknown length
* 29930 31543: contig of 1614 bp in length
*      gap of unknown length
* 31544 33044: contig of 1501 bp in length
*      gap of unknown length
* 33045 35128: contig of 2084 bp in length
*      gap of unknown length
* 35129 37580: contig of 2452 bp in length
*      gap of unknown length
* 37581 38859: contig of 1279 bp in length
*      gap of unknown length
* 38860 40904: contig of 2045 bp in length
*      gap of unknown length
* 40905 43040: contig of 2136 bp in length
*      gap of unknown length
* 43041 44561: contig of 1521 bp in length
*      gap of unknown length
* 44562 46554: contig of 1993 bp in length
*      gap of unknown length
* 46555 48508: contig of 1954 bp in length
*      gap of unknown length
* 48509 49737: contig of 1229 bp in length
*      gap of unknown length
* 49738 51391: contig of 1654 bp in length
*      gap of unknown length
* 51392 52987: contig of 1596 bp in length
*      gap of unknown length
* 52988 55258: contig of 2271 bp in length
*      gap of unknown length
* 55259 58043: contig of 2785 bp in length
*      gap of unknown length
* 58044 59391: contig of 1348 bp in length
*      gap of unknown length
* 59392 62496: contig of 3105 bp in length
*      gap of unknown length
* 62497 65618: contig of 3122 bp in length
*      gap of unknown length
* 65619 67648: contig of 2030 bp in length
*      gap of unknown length
* 67649 70075: contig of 2427 bp in length
*      gap of unknown length
* 70076 73261: contig of 3186 bp in length
*      gap of unknown length
* 73262 76990: contig of 3629 bp in length
*      gap of unknown length
* 76991 80022: contig of 3132 bp in length
*      gap of unknown length
* 80023 89215: contig of 9193 bp in length
*      gap of unknown length
* 89216 98779: contig of 9564 bp in length
*      gap of unknown length
* 98780 132048: contig of 33269 bp in length
*      gap of unknown length
* 132049 190353: contig of 58305 bp in length
*      gap of unknown length
* 190354 245472: contig of 55119 bp in length.
Location/Qualifiers
1..245472
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-61L4"

```

FEATURES source

```

BASE COUNT 64057 a 60702 c 58925 g 61440 t 348 others
ORIGIN
alignment_scores:
    Quality: 104.50      Length: 96
    Ratio: 2.133        Gaps: 4
    Percent Similarity: 51.042    Percent Identity: 36.458
alignment_block:
US-09-135-238B-2_COPY_273_390 x AC027687 ..
Align seg 1/1 to: AC027687 from: 1 to: 245472
      2 ArgAlaValGluArgArg..... 7
      172086 CGCGCGCGGAGCGAGACTCCCGGCTCTCCCTCCCTCGGCTC 172135
      8 .....LysAlaLeuSerArgArgAlaArgAlaArgLeuAlav 19
      172136 GTGACAACGACGCCCGGCTGTAGCGGCGCGCGCGCGGTCG 172185
      19 alArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerProArg 35
      172186 ACCGGCTGAGCGGAGAGGAGC.....CGGCCTCGGGGCTCGCCCGC 172229
      36 ProArgSerGlnAsnAsnIleTyrSerAlaCysPro.....ArgAr 49
      172230 CCCCCG.....GTCTTGACTCCGCGGCCCTCTCTCGGCG 172267
      49 gAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlaProValProGlyP 66
      172268 CTTCCGTTACGCCCGTTGTCGCGGCGGCGCGCGGCCCATCCCGAGC 172317
      66 roGlyAlaProLeuProAlaProLeuGlnValSer 78
      172318 CGAGGCGCTGTGTCGCCCGCTCGCAGCGCCAGTGAGC 172355
seq_name: gb_bal:SCF43A
seq_documentation_block:
LOCUS SCF43A 35437 bp DNA BCT 13-JUL-1999
DEFINITION Streptomyces coelicolor cosmid F43A.
ACCESSION AL096837
VERSION AL096837.1 GI:5459214
KEYWORDS ABC transporter; alkaline phosphatase; ATP-binding protein;
deoxyribodipyrimidine photolyase; ferredoxin; gntR-family;
iron-sulfur protein; isomerase; lyase; oxidoreductase;
penicillin-binding protein; phr; rhamnose kinase; serine protease;
threonine dehydratase; transcriptional regulator; transmembrane
transport protein; transport protein.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2)
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 35437)
AUTHORS Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
REFERENCE 2 (bases 1 to 35437)
AUTHORS Seeger,K. and Harris,D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 35437)
AUTHORS James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK

```

COMMENT	Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC787.01c, SC (S. coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl . CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid F43A lies between cosmids F43 and M2 on the AseI-F region of the chromosome.
FEATURES	Location/Qualifiers 1..35437 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid F43A" complement(1..312) /gene="SCF43A.01c" complement(<1..312) /gene="SCF43A.01c" /note="SCF43A.01c, probable ABC transporter ATP-binding protein, partial CDS, len: >104 aa; similar to many e.g. SW:ARAG_ECOLI (EMBL:X06091), araG, Escherichia coli L-arabinose transport ATP-binding protein (504 aa), fasta scores: opt: 305 z-score: 368.0 E(): 3.6e-13, 45.8% identity in 96 aa overlap. Similar to many others from S.coelicolor e.g. TR:O50504 (EMBL:AL009199) S.coelicolor probable ABC-transport system ATP binding protein (260 aa) (55.2% identity in 96 aa overlap). Contains pfam match to PF00005 ABC_tran, ABC transporter and PS00017 ATP/GTP-binding site motif A (P-loop)" /codon_start=1 /transl_table=11 /label="SCF43A.01c" /product="putative ABC transporter ATP-binding protein" /protein_id="CAB48888.1" /db_xref="GI:5459215" /translation="MTHFSRDTGSPVLAIRDVSKFGAVRALPDVSLFGEVHALA GENGAGKSTLIKTLAGVHRPDAGHVLIDGAPVVFHGGDANDAGIAVIITGFLFPOL SI"
misc_feature	1..101 /note="overlap with Streptomyces coelicolor cosmid F43 from 22296 to 22396" complement(157..180) /gene="SCF43A.01c"
misc_feature	/note="PS00017 ATP/GTP-binding site motif A (P-loop)" 577..581 /note="possible RBS"
RBS	589..1749
gene	/gene="SCF43A.02"
misc_feature	1868..3907 /note="SCF43A.02, possible sugar isomerase, len: 386 aa; weakly similar to many xylose isomerases (divergent at N- and C-termini) e.g. SW:XYLA_AMPSP (EMBL:M15050), xyIA, Ampullaria sp. xylose isomerase (394 aa), fasta scores: opt: 189 z-score: 229.6 E(): 1.8e-05, 26.9% identity in 219 aa overlap" /codon_start=1 /transl_table=11 /label="SCF43A.02" /product="putative sugar isomerase" /protein_id="CAB48889.1" /db_xref="GI:5459216" /translation="MTELAAYKAALKTQAVETPSWAYGNSGTRFKVFAQGVPRDPFE KIDDAAVHEFTGAPTVALHPIWDVDEYAALAAEKGRVIGAINSTFQDDDYR LGSTCHPDAAVRKAVRKHLLCEVDIMDATGSDRLKWFADGTNPGQDDIRSDRLA EGLAEVTERGEGQRMLLFYLKLPFAFYTTDVMGTAYAHCLKLGEKAAOVVDYTGHH APTGNISFIVATLREGKIGGFDSRFYADDDLMVGAADPFQFLFRIMYVVRGGFT SDVAFMLDQCHNLEAKIPALIRSNVYQETAKALIVDTALAEQAAGDVLENAVLT MDATNYDRLREVRESGLDPEPMKAYRSCGKAEKVVAERIGGQQAQNGA"
gene	1868..3907 /gene="SCF43A.03"
CDS	1868..3907 /gene="SCF43A.03" /note="SCF43A.03, probable oxidoreductase, len: 679 aa; similar to many members of the short-chain dehydrogenase/reductase family e.g. SW:YUXG_BACSU (EMBL:J93938) Bacillus subtilis hypothetical oxidoreductase (689 aa), fasta scores: opt: 1727 z-score: 1840.8 E(): 0, 40.3% identity in 677 aa overlap. C-terminal half is similar to acetoin dehydrogenases e.g. SW:BUDC_KLEPN (EMBL:D86412), budC, Klebsiella pneumoniae acetoindiacetyl reductase (256 aa) (35.7% identity in 258 aa overlap). C-terminal half is similar to putative dehydrogenases from S.coelicolor e.g. TR:O88068 (EMBL:AL031541) S.coelicolor probable dehydrogenase (260 aa) (34.3% identity in 268 aa overlap). Contains probable coiled-coil from 307 to 334 (28 residues) (Max score: 1.492, probability 0.94). Contains pfam match to entry PF00106 adh_short, short chain dehydrogenase" /codon_start=1 /transl_table=11 /label="SCF43A.03" /product="putative oxidoreductase" /protein_id="CAB48890.1" /db_xref="GI:5459217" /translation="MGIHPSAALLARSHRIGADPRNTNAGGNASAKGTGTPVTGG EVELMMVKGSGDGLTGTAGLAVLRDMRLKDVYPGVEHEDEMVAADFYLHGKG GAAPSIDTAMHGLVEAAHYDHLHPDSDGIALACADGEKLTAECDTVVWVWRPFG QLGDLIAVKEANPQAVGCVLGGHGTITAGDSTEECERNLHIITAEATLAEGRKAE PPGPVIYAAALPEAERRERAAALAPYVVRGLASRDRAQVGHFTDADVDFLARAEHP RLAAUGTSCDPHFTRTKVPRLVDVAPTALEAVARLAKELHAAVREYAAVYERHAE PPSPMRGADPAIVLPVGVGMESFGCKQKQARVAGFEYVNAINMARGAEAVTPAIE ESEKPRIDYMALEAKLRMPKPKELATRLVLTGAGSGIGKAIARLIVDEGACVVA DLNAENAAVAEEGGDKXAVAVTVDVTSQQTAAPFAALAFGGVDLVVNNAGLSI SXPLETSKADMDLIQHDIMARGSLVRSREARVMTAEQELGGDVIITIAKSNKAAVAGPN IYATSKAQDQHQVRLAAELGEHGRVNGVNFDPVYVSGSIFAGGMAKRAAYGVYP EKLKGEFTAQRTLLKREVLPEHVANAVFALTGGDLTHTTGLHVPVDAGVAAFLR"
misc_feature	3128..3685 /gene="SCF43A.03" /note="pfam match to entry PF00106 adh_short, short chain dehydrogenase, score 197.80, E-value 1.7e-55" 3904..5358 /gene="SCF43A.04"
gene	3904..5358 /gene="SCF43A.04"
CDS	3904..5358 /gene="SCF43A.04" /note="SCF43A.04, probable rhamnose kinase, len: 484 aa; similar to rhamnose kinases e.g. SW:RHAB_ECOLI (EMBL:L19201), rhaB, Escherichia coli rhamnulokinase (489 aa), fasta scores: opt: 1136 z-score: 1280.7 E(): 0, 41.5% identity in 488 aa overlap. Also similar to many other carbohydrate kinases e.g. SW:FUCK_ECOLI (EMBL:X15025),

CDS
 4505..4762
 /note="acyl carrier protein (AA 1-85)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA34266.1"
 /db_xref="GI:47981"
 /db_xref="SWISS-PROT:P12885"
 /translation="MARLTDLGLRIILVACAGEDGVDLSDILDITFEELGYDSLAL
 MESASRIEELGVALADGDINIELIPRVLLDLVNGAQAEA"
 4882..5856
 /note="URF (AA 1-324)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA34267.1"
 /db_xref="GI:47982"
 /db_xref="SWISS-PROT:P16560"
 /translation="WQPAATPVSLPSTVHRSEHTVTVAAPPEALYALVADVTRPPA
 VFETVHVRLEAGESTEFELVAEINGEIAHWRSSRVLDPRYIVSFROEHSRPPT
 SMGGVFLRLADGRTEIYVLRHFTVADDPAAVRIEALDRNSARELGALAALET
 GHPVDELVSFTDTLQCAARADATFYVERAERWAEELPHVACQCADRAGTGLEQMLE
 MDVTADGSHHTTFSRACIAPEWTAYNEORTPLRVSGHSGEWTFAOTPGPVATARH
 TVAVDPGSEITVELGDPATLADARAYLRALGNSLATLRHAAEAQRA"
 BASE COUNT 903 a 2666 c 2352 g 872 t
 ORIGIN

 alignment_scores:
 Quality: 103.50 Length: 69
 Ratio: 2.797 Gaps: 3
 Percent Similarity: 53.623 Percent Identity: 37.681

 alignment_block:
 US-09-135-238B-2_COPY_273_390 x SVGRA/rev ..

 Align seg 1/1 to reverse of: SVGRA from: 1 to: 6793

 23 AlaLeuGluSerSerGlnArgProArgGlySerProArgProArgSerG1 39
 ||| :::::||||| ||||| ::||| |||||
 1870 GCCCGCGGTACCTACCTACGTACACCCCGGCACACTCCCGGCC..... 1829

 39 nAsnAsnIleTyrSerAlaCysProArgAlaArgGlyAlaAspAla 56
 ||| ||| ||| ||||| |||
 1828TGCTCAGAGCGGCTCGCGCGCGATGGGCG 1798

 56 laGlyThrGlyGluAlaProValProGlyProGlyAlaProLeuPro... 71
 ||| |||:::||||| ||| ||| ::|||::|||
 1797 GCGGAGAGGCGCATGGGTGCTGCCCTGCCCGCAACCCCATCCGCTC 1748

 72ProAlaProLeuGlnValSerGluSerProTppLeuH1 84
 ||| ||||| ||||| ||||| |||||
 1747 TCCATCGGGCTTCACGCCCGCTC.....GAGGCCCTTGCGACG 1707
 ::|||
 84 sAlaPro 86
 1706 CTCCTCCC 1700

 seq_name: gb_ro:BTPRMTPN2

 seq_documentation_block:
 LOCUS BTPRMTPN2 13187 bp DNA ROD 11-DEC-1996
 DEFINITION R.norvegicus PRM1, PRM2, PRM3 and TNP2 genes.
 ACCESSION Z46939
 VERSION Z46939.1 GI:1359527
 KEYWORDS PRM1 gene; PRM2 gene; PRM3 gene; protamine; TNP2 gene; transition
 protein 2.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

[illegible]

OM of: US-09-135-238B-2_COPY_18_253 to: N_Geneseq_36.* out_format : pfs
Date: Sep 12, 2000 6:55 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q/cn2/_USPTO.spool/US09135238/runat_24082000_091324_17918/app_query.fasta_1.1282
-DB=N_Geneseq_36 -QPMF=fastap -SUFFIX=p2n.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-CGAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonsum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=100000
-USER=US09135238 @CGN1_1_123 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPVY
-WAIT -THREADS=1

Search information block:

Query: US-09-135-238B-2_COPY_18_253
Query length: 236
Database: N_Geneseq_36.*
Database sequences: 311585
Database length: 125096042
Search time (sec): 121.040000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:X28178	+ 1254.00	1961.26	1.2e-101	2040	Human PIGRL-1 coding sequence.
N_Geneseq_36:X28179	+ 1029.50	1613.23	2.9e-82	1047	Human PIGRL-1 coding sequence.
N_Geneseq_36:X12190	+ 197.00	287.40	2.1e-08	3095	Mouse poly-Ig receptor protein
N_Geneseq_36:X15422	+ 188.00	250.87	2.3e-06	2932	Mouse poly-Ig receptor protein
N_Geneseq_36:X131291	+ 185.50	268.69	2.3e-07	3269	Rat poly-immunoglobulin receptor
N_Geneseq_36:X07407	+ 180.50	266.49	3.0e-07	1839	Human secretory immunoglobulin
N_Geneseq_36:X30857	+ 180.50	285.99	3.2e-07	1936	Secretory component DNA fragment
N_Geneseq_36:X30858	+ 180.50	265.51	3.5e-07	2031	Secretory component DNA fragment
N_Geneseq_36:X311287	+ 180.50	261.92	5.5e-07	2919	Human poly-immunoglobulin receptor
N_Geneseq_36:X31289	+ 179.50	258.18	8.8e-07	3630	Bovine poly-immunoglobulin receptor
N_Geneseq_36:X20383	+ 175.50	238.22	8.8e-07	1911	cDNA for human immunity related
N_Geneseq_36:X14498	+ 166.50	244.18	5.3e-06	1876	Clone 1 for truncated poly-Ig
N_Geneseq_36:X11287	+ 166.50	237.95	1.2e-05	3517	Rabbit poly-immunoglobulin receptor
N_Geneseq_36:X14499	+ 145.00	210.22	0.0004	1876	Clone 2 for truncated poly-Ig
N_Geneseq_36:X28205	+ 119.50	174.90	0.0385	1137	Human PIGR-2 coding sequence.
N_Geneseq_36:X96035	+ 117.50	167.21	0.1031	1795	Human kidney injury related molecule
N_Geneseq_36:X87689	+ 116.00	176.30	0.0321	565	EST clone EN260. New polynucleotide
N_Geneseq_36:X96033	+ 111.50	156.26	0.4203	2084	Rat kidney injury related molecule
N_Geneseq_36:X96032	+ 111.50	154.19	0.5476	2566	Rat kidney injury related molecule
N_Geneseq_36:X34234	+ 107.50	156.70	0.3969	1053	Human secreted protein gene 16
N_Geneseq_36:X34235	+ 107.50	156.50	0.4075	1075	Human secreted protein gene 16
N_Geneseq_36:X34169	+ 106.50	154.65	0.5168	1105	Human secreted protein gene 16
N_Geneseq_36:X18694	+ 104.50	131.05	10.66	8690	Plasmid pD17-hum14.H1. Inhibitor
N_Geneseq_36:X41432	+ 104.00	148.16	1.19	1427	Plasmid Hu19Hpcpc encoding an
N_Geneseq_36:X41429	+ 103.00	146.58	1.45	1427	Plasmid Hu19Hpcpc encoding an
N_Geneseq_36:X41431	+ 103.00	146.58	1.45	1427	Plasmid Hu19Hpcpc encoding an
N_Geneseq_36:X18598	+ 102.50	128.32	15.13	8321	Plasmid expressing hBR95-A. In
N_Geneseq_36:X15932	+ 102.00	129.89	12.37	6557	Anti-IgE VH expression vector.
N_Geneseq_36:X38716	+ 101.50	136.24	5.48	3190	CD22-beta gene. Inhibition of
N_Geneseq_36:X62510	+ 101.00	143.40	2.19	1431	Primate anti-human B7.1 and
N_Geneseq_36:X35485	+ 101.00	143.40	2.19	1431	Macaque primatized 7C10 heavy
N_Geneseq_36:X03509	+ 100.50	141.39	2.83	1617	Antibody D heavy chain. Prodn.
N_Geneseq_36:X41427	+ 100.00	127.15	17.57	6284	Plasmid Hu19AHpcpc encoding an
N_Geneseq_36:X60665	+ 99.50	130.46	11.50	4157	Human CD100 antigen cDNA. Nucl
N_Geneseq_36:X24074	+ 99.00	144.68	1.86	914	Human EST R2810 DNA. Complexes
N_Geneseq_36:X66631	+ 99.00	137.59	4.61	1869	Human immunoglobulin heavy chain
N_Geneseq_36:X080470	+ 98.50	145.41	1.69	784	Anti-phenylazolonone antibody c
N_Geneseq_36:X80580	+ 98.50	142.42	2.48	1060	Kidney injury associated molecul
N_Geneseq_36:X07474	+ 98.50	125.03	23.06	6127	Musculus anti-IgE antibody
N_Geneseq_36:X049834	+ 97.00	137.39	4.72	1386	Anti-HIV-1 recombinant antibody

N_Geneseq_36:Q12632	+	97.00	132.10	9.31	2364	! CDA-specific CDR-grafted he
N_Geneseq_36:Q49944	+	96.50	135.33	6.15	1576	! Human anti-HBs heavy chain.
N_Geneseq_36:X52228	+	96.50	131.63	9.90	2290	! Protein PRO27 cDNA clone D
N_Geneseq_36:V38929	+	96.00	132.49	8.86	1938	! A33 chimeric receptor DNA.
N_Geneseq_36:N81713	+	96.00	127.69	16.40	3146	! Clone contg. thrc gene. Thr

seq_name: N_Geneseq_36:X28178

seq_documentation_block:

ID X28178 standard; cDNA; 2040 BP.
AC X28178;
DT 16-JUN-1999 (first entry)
DE Human PIGRL-1 coding sequence.
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;
KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;
KW diagnosis; therapy; ss.
OS Homo sapiens.
PN EP-905238-A2.
PD 31-MAR-1999.
PE 14-AUG-1998; 306487.
PR 30-OCT-1997; US-961564.
PR 25-AUG-1997; US-056935.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Sweet RW, Truneh A, Wu S;
DR WPI: 99-192666/17.
DR P-PSDB: Y05001.
PT New polypeptides encoding PIGRL-1 useful for treating diseases such
as X-linked Severe Combined Immunodeficiency
PS Claim 4; Page 18-19; 26pp; English.
CC This sequence encodes the human PIGRL-1 protein of the invention.
CC Autoimmune diseases involving altered expression or activity of PIGRL-1
CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined
CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be
CC diagnosed or susceptibility to them predicted by: (1) determining whether
CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or
CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.
CC Patients deficient in PIGRL-1 can be treated by administering either the
CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.
CC Patients with excessive expression or activity of PIGRL-1 can be treated
CC by administering an antagonist of PIGRL-1, an antisense nucleic acid
CC molecule which inhibits the expression of PIGRL-1 or administering
CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can
CC be used to identify its agonists by contacting a cell expressing PIGRL-1
CC with a candidate compound in the presence of a signal system and noting
CC the candidate as an agonist if a signal is produced. The same method can
CC be used to identify antagonists of PIGRL-1 but the presence of an
CC antagonist is indicated by a decrease in production of the signal.
CC Antibodies against PIGRL-1 may be used to isolate or identify clones
CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to
CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.
CC This information may then be correlated with the incidence of autoimmune
CC disease in those patients to identify whether the mutation causes the
CC disease.

Sequence 2040 BP; 478 A; 616 C; 494 G; 452 T;

alignment_scores:

Quality: 1254.00 Length: 236
Ratio: 5.314 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-135-238B-2_COPY_18_253 x X28178 ..
Align seg 1/1 to: X28178 from: 1 to: 2040
1 Arg1LeuProGluValLysValGluGlyLeuGlySerValth 17
|||||
197 AGGATCTCCCAAGTAAGGTAGAGGGAGCTGGCGGATCATGTTAC 246
|||||
17 r1LeLysCysProLeuProGluMethHisValArgIleTyrLeuCysArg 34
|||||
247 CATCAAGTGGCCACTTCTCTCAATGCATGTGAGGATATATCTGTGCGGG 296
|||||

```
34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
|||||
297 AGATGGCTGGATCTGGAACATGTGTACCGTGTATCCACCACTTC 346
|||||
51 IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLys 67
|||||
347 ATCAAGGCAGAAATACAGGGCCGAGTTACTCTGAAGCAATACCCACGAA 396
|||||
67 sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84
|||||
397 GAATCTGTCTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGAG 446
|||||
84 alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
|||||
447 TCTATGCTCGGAGCGGCATGAACACAGACCGGGGAAGACCCAGAA 496
|||||
101 ValThrLeuAsnValHisSerGluTyrGluProSerTrpGluGlnPr 117
|||||
497 GTCACCTGAATGTCACAGTGAATACGAGCCATCATGGGAAGACGCC 546
|||||
117 oMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMetP 134
|||||
547 AATGCTGAGACTCCAAAATGGTTTCATCTGCCCTATTGTTCAGATGC 596
|||||
134 roAlaTyrAlaSerSerSerLysPheValThrArgValThrThrProAla 150
|||||
597 CTGCATATGCCAGTCTCTCCAAATTCGTAAACAGAGTTACACACAGCT 646
|||||
151 GlnArgGlyLysValProProValHisHisSerSerProThrThrGlnIl 167
|||||
647 CAAAGGGCAGGTGCTCTCCAGTTCACACTCTCTCCCCACACCCCAAT 696
|||||
167 eThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLysP 184
|||||
697 CACCACCGCCCTCGAGTGTCCAGAGCATCTTCAGTAGCAGGTGACAA 746
|||||
184 roArgThrPheLeuProSerThrThrAlaSerLysLysSerAlaLeuGlu 200
|||||
747 CCCGAACCTCTCTGCCATCCACTACAGCTCAAAAATCTCAGCTCTGGAG 796
|||||
201 GlyLeuLeuLysProGlnThrProSerTyrAsnHisHisThrArgLeuHi 217
|||||
797 GGGCTGCTCAGCCCCAGAGCCGACCTACACCCACACAGGCTGCA 846
|||||
217 sArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGlyGlnG 234
|||||
847 CAGGCAGAGAGCACTGGACTATGGCTCACAGTCTGGGAGGGAAGGCCAAG 896
|||||
234 lyPheHis 236
|||||
897 GATTCAC 904

seq_name: N_Geneseq_36.X28179
seq_documentation_block:
ID X28179 standard; cDNA; 1047 BP.
AC X28179;
AT
DT 16-JUN-1999 (first entry)
DE Human PIGRL-1 coding sequence fragment.
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;
KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;
KW diagnosis; therapy; ss.
OS Homo sapiens.
PN EP-905238-A2.
PD 31-MAR-1999.
PF 14-AUG-1998; 306487.
PR 30-OCT-1997; US-961564.
PR 25-AUG-1997; US-056935.
PA (SMK ) SMITHKLINE BEECHAM CORP.
PI Sweet RW, Truneh A, Wu S;
DR WPI; 99-192666/17.
DR P-PSDB; Y05002.
```

```
PT New polypeptides encoding PIGRL-1 useful for treating diseases such
as X-linked Severe Combined Immunodeficiency
PS Disclosure: Page 8; 26pp; English.
CC This sequence encodes the human PIGRL-1 protein of the invention.
CC Autoimmune diseases involving altered expression or activity of PIGRL-1
may include hyper-IgM immunodeficiency (HIM), X-linked Severe Combined
CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be
CC diagnosed or susceptibility to them predicted by: (1) determining whether
CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or
CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.
CC Patients deficient in PIGRL-1 can be treated by administering either the
CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.
CC Patients with excessive expression or activity of PIGRL-1 can be treated
CC by administering an antagonist of PIGRL-1, an antisense nucleic acid
CC molecule which inhibits the expression of PIGRL-1 or administering
CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can
CC be used to identify its agonists by contacting a cell expressing PIGRL-1
CC with a candidate compound in the presence of a signal system and noting
CC the candidate as an agonist if a signal is produced. The same method can
CC be used to identify antagonists of PIGRL-1 but the presence of an
CC antagonist is indicated by a decrease in production of the signal.
CC Antibodies against PIGRL-1 may be used to isolate or identify clones
CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to
CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.
CC This information may then be correlated with the incidence of autoimmune
CC disease in those patients to identify whether the mutation causes the
CC disease.
SQ Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T;
```

```
alignment_scores:
Quality: 1029.50 Length: 245
Ratio: 4.457 Gaps: 10
Percent Similarity: 94.286 Percent Identity: 92.245

alignment_block:
US-09-135-238B-2_COPY_18_253 x X28179 ..

Align seg 1/1 to: X28179 from: 1 to: 1047

1 ArgIleLeuProGluValLysValGluGlyGluLeuGlyGlySerValTh 17
|||||
201 AGGATCTCCAGAGTAAGTAAAGTAGAGGGGAGCTGGCGGATCAGTTAC 250
|||||
17 rIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArgG 34
|||||
251 CATCAAGTCCCACTTCCTGAAATGTCATGTGAGGATATATCTGTGCGGG 300
|||||
34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
|||||
301 AGATGGCTGGATCTGGAACATGTGTACCGTGTATCCACCACTTC 350
|||||
51 IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67
|||||
351 ATCAAGGCAGAAATACAGGGCCGAGTTACTCTGAAGCAATACCCACGAA 400
|||||
67 sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84
|||||
401 GAATCTGTCTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGAG 450
|||||
84 alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
|||||
451 TCTATGCTCGGAGCGGCATGAACACAGACCGGGGAAGACCCAGAAA 499
|||||
101 ValThrLeuAsnValHisSerGluTyrGluProSerTrpGluGlnPr 117
|||||
500 GTCACCTGAATGTCACAGTGAATACGAGCCATCATGGGAAGACGCC 549
|||||
117 oMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMetP 134
|||||
550 AATGCTGAGACTCCAAAATGGTTTCATCTGCCCTATTGTTCAGATGC 599
|||||
134 roAlaTyrAlaSerSerSerLysPheValThr..ArgValThrThrProA 150
```

```

|||||.....|
600 CTGCATATGCGGTCTCTCCACATTCGTAACCCGAGATTACACACCAG 649
150 la.GlnArgGlyLysValProProValHisHisSerSerProThrThrG1 166
|||||.....|
650 CTTCAAGGGGGGCAAGGTCCCTCCAGTTCACCACTCTCCCTCCACACCCA 699
166 nile.ThrHisArgPro.ArgValSerArgAlaSerSerValAlaGlyAs 182
|||||.....|
700 AATTCACCCACCCGCTTCGAGTGTNCAGAGCATCTTCAGTACAGGTGA 749
182 pLysProArg.ThrPheLeuProSerThrThrAlaSerLysLysSerAla 198
|||||.....|
750 CAAGCCCGCAACTTCTCCCTCCACTCCACTACAGCTCAAAAATCTCAGCT 799
199 LeuGluGlyLeuLys.....ProGlnThrProSerTyrAsnHisHisTh 214
|||||.....|
800 CTGAAGGGGTCTTCAAGCCCGCAAGAGGCCAGCTACAAACAN.CACAC 848
214 rArgLeuHisArgGlnArgAlaLeuAsp...Tyr.GlySerGlnSerGly 229
|||||.....|
849 CAGGCTGCACAGCAGCAGACACTGGATCTTATGGNTCAGACTCGG 898
230 ArgGluGly.....GlnGlyPhe 235
|||||.....|
899 ...GAGGGGAANGNCCAAGGATT 919

```

seq_name: N_Geneseq_36:T31290

seq_documentation_block:

ID T31290 standard; cDNA; 3095 BP.

AC T31290;
 DE 24-FEB-1997 (first entry)
 DT Mouse poly-immunoglobulin receptor, cDNA.
 KW Mouse; immunoglobulin; receptor; protection protein; mutants;
 KW heavy chain; antigen binding domain; protection; pathogen;
 KW mucosal; environment; gastrointestinal; passive; immunisation;
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
 KW poly; sorbinus; murine; ss.
 OS Mus musculus.

FH Key Location/Qualifiers

FT cds 85..2400

ET /*tag= a

PN W09621012-A1.

PD 11-JUL-1996.

PF 27-DEC-1995; U16889.

PR 30-DEC-1994; US-367395.

PR 04-MAY-1995; US-434000.

PA (PLAN-) PLANT BIOTECHNOLOGY INC.

PA (UNKE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.

PA (PLAN-) PLANET BIOTECHNOLOGY INC.

PI Hiatt AC, Lehner T, Ma JKC;

DR WPI; 96-333987/33.

DR P-PSDB; W03180.

PT Immunoglobulin and protection protein complex and its prodn. in
 PT plants - useful for passive immunisation against mucosal antigens,
 PT esp. against S. mutans and S. sorbinus to prevent dental caries,
 PS Disclosure; pages 117-121; 152pp; English.

CC The present sequence encodes the mouse poly-immunoglobulin (Ig)
 CC receptor, a portion of which corresp. to residues 1-627, pref.
 CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
 CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).

CC The Ig of the invention comprises a PP as above in association with
 CC an Ig derived heavy chain, having at least a portion of an antigen
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
 CC gastrointestinal, environments, therefore enhancing its

CC effectiveness in passively immunising animals against mucosal
 CC pathogens. The Ag binding domain is specifically derived from the
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
 CC S. sorbinus serotypes d and g.

SQ Sequence 3095 BP; 861 A; 796 C; 784 G; 654 T;

alignment_scores:

Quality: 197.00 Length: 148

Ratio: 2.165 Gaps: 5

Percent Similarity: 61.486 Percent Identity: 32.432

alignment_block:

US-09-135-238B-2_COPY_18_253 x T31290 ..

Align seg 1/1 to: T31290 from: 1 to: 3095

```

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
||| |||||.....|
181 GCGGACTCTGTTTCCATCAGCTGCTACTACCAGACACACTCTGTCAACCG 230
25 ThisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||| |||||.....|
231 GCACACCGGAAATACTGGTCCGACCAAGGAGCC....AGCGGCATGTGCA 277
42 lyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 58
||| |||||.....|
278 CAAGCCTCATCTCTCAATGGCTACCTCTCCAAGGAGTATTCAGGCAGA 327
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
||| |||||.....|
328 GCCAACCCTCATCACTCCAGAGAACACACACATTTGTGATTACATGTA 377
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92
||| |||||.....|
378 GCAGCTCACCAGGACGACACTGGGAGCTACAAAGTGTGGCTGGGT...A 424
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
||| |||||.....|
425 CCAGTAACCCAGGCGCTCTCTTCGATGTCAGCCTGGAGGTC..... 465
109 TyrGluProSerTrpGluGlnProMetProGluThrProLysTrpPh 125
||| |||||.....|
466 .....AGCCAGGTCTCTGAGTTCGCCGAGTACAC 494
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 142
||| |||||.....|
495 CCAGCTC.....TACACAAAGGACATAGGCA 520
142 heValThrArgValThrThrProAlaGlnArgGlyLysValPro 156
||| |||||.....|
521 GAAATGTGACCATGATGCCCTTCAAAAGGAGGAATGTCC 564

```

seq_name: N_Geneseq_36:V15422

seq_documentation_block:

ID V15422 standard; DNA; 29392 BP.

AC V15422;

DT 11-JUN-1998 (first entry)

DE Mouse poly Ig receptor protein gene.

KW Mouse; poly Ig receptor protein; pIgR protein; pIg deficiency;

KW knockout mouse; disease model; ds.

OS Mus sp.

PN J10057066-A.

PD 03-MAR-1998.

PF 19-AUG-1996; 217154.

PR 19-AUG-1996; JP-217154.

PA (HONS) YAKULT HONSHA KK.

DR WPI; 98-254323/23.

PT Mouse pIg receptor protein gene - used for preparing gene knockout
 PT mice, useful for study of human poly Ig receptor protein deficiency
 PS Claim 1; Page 4-14; 18pp; Japanese.

CC The present sequence represents the mouse poly Ig receptor protein
 CC gene, which has a 29392 bp sequence. The new gene can be used to
 CC produce a gene knockout mouse, useful as a disease model of human
 CC poly Ig receptor protein deficiency.

SQ Sequence 29392 BP; 8318 A; 6747 C; 6514 G; 7813 T;

OM of: US-09-135-238B-2 to: N_Geneseq_36:* out_format : pfs

Date: Sep 12, 2000 6:55 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09135238/runat_24082000_091324.17918/app_query.fasta_1.1282
-DB=N_Geneseq_36 -QFMT=fastap -SURFIX=p2n.rng -GAPOP=12.000
-CAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=100000
-USER=US09135238 -CGN1_1.123 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-135-238B-2
Database: N_Geneseq_36:*
Query length: 390
Database sequences: 311585
Database length: 125096042
Search time (sec): 121.040000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
N_Geneseq_36:X28178	+ 2047.00	1229.34	5.3e-111	2040	Human PIGRL-1 coding sequence.
N_Geneseq_36:X28179	+ 1125.00	1174.09	8.5e-58	1047	Human PIGRL-1 coding sequence.
N_Geneseq_36:T31220	+ 188.00	199.21	0.0017	3095	Mouse poly-immunoglobulin rece
N_Geneseq_36:T31221	+ 198.00	170.27	0.0697	29392	Mouse poly Ig receptor protei
N_Geneseq_36:T31222	+ 187.50	192.24	0.0042	1911	CDNA for human immunity relat
N_Geneseq_36:T31223	+ 185.50	185.74	0.0096	3269	Rat poly-immunoglobulin recept
N_Geneseq_36:T31224	+ 185.00	189.53	0.0059	1839	Secretory component DNA fragme
N_Geneseq_36:T31225	+ 182.00	186.82	0.0083	1939	Human secretory immunoglobulin
N_Geneseq_36:T31226	+ 182.00	186.01	0.0093	2031	Secretory component DNA fragme
N_Geneseq_36:T31227	+ 182.00	183.02	0.0136	2919	Human poly-immunoglobulin rece
N_Geneseq_36:T31228	+ 182.00	181.23	0.0171	3630	Bovine poly-immunoglobulin rec
N_Geneseq_36:T31229	+ 171.00	174.33	0.0414	2084	Rat kidney injury related mole
N_Geneseq_36:T31230	+ 171.00	172.62	0.0515	2566	Rat kidney injury related mole
N_Geneseq_36:T31231	+ 166.50	170.51	0.0676	1876	Clone 1 for truncated poly Ig
N_Geneseq_36:T31232	+ 166.50	165.34	0.1312	3517	Rabbit poly-immunoglobulin rec
N_Geneseq_36:T31233	+ 129.50	167.34	0.1015	1137	Human PIGR-2 coding sequence.
N_Geneseq_36:T31234	+ 145.00	148.11	1.20	1876	Clone 2 for truncated poly Ig
N_Geneseq_36:T31235	+ 143.50	133.93	7.37	8690	Plasmid pDI7-hm14.H1. Inhibit
N_Geneseq_36:T31236	+ 142.00	150.90	0.8361	914	Human EST R2810 DNA. Complexes
N_Geneseq_36:T31237	+ 139.50	130.12	12.02	8321	Plasmid expressing hBR96-A. In
N_Geneseq_36:T31238	+ 138.50	146.63	1.44	985	Mycobacterium tuberculosis anti
N_Geneseq_36:T31239	+ 138.50	146.63	1.44	985	M. tuberculosis immunogenic pol
N_Geneseq_36:T31240	+ 136.50	149.12	1.05	565	EST clone EN260. New polynucle
N_Geneseq_36:T31241	+ 136.50	140.14	3.32	1683	Human dUTPase gene 5' region.
N_Geneseq_36:T31242	+ 136.00	123.79	27.06	11529	Plasmid pAH4602. Antibody con
N_Geneseq_36:T31243	+ 135.00	123.35	28.60	10704	Plasmid pAH4625. Antibody con
N_Geneseq_36:T31244	+ 131.00	118.16	55.71	12132	Plasmid pAH4807. Antibody con
N_Geneseq_36:T31245	+ 130.00	133.70	7.59	1617	Antibody D heavy chain. Prodn.
N_Geneseq_36:T31246	+ 129.50	134.18	7.13	1431	Primate anti-human B7.1 ant
N_Geneseq_36:T31247	+ 129.50	134.18	7.13	1431	Macaque primatized 7C10 heavy
N_Geneseq_36:T31248	+ 127.00	128.05	15.66	2197	IS1096 transposon. Leucine aux
N_Geneseq_36:T31249	+ 126.50	120.88	39.26	4266	Plasmid pHC65 nucleotide sequ
N_Geneseq_36:T31250	+ 126.50	114.03	94.63	11336	Approximate nucleotide seque
N_Geneseq_36:T31251	+ 125.50	129.22	13.48	1576	Human anti-HBs heavy chain. Hu
N_Geneseq_36:T31252	+ 125.50	122.89	30.35	3400	3F4 human G2/G4 chimeric anti
N_Geneseq_36:T31253	+ 125.50	116.64	67.64	7266	Human calcium channel a1B sub
N_Geneseq_36:T31254	+ 124.50	116.45	69.37	6557	Anti-IgE VH expression vector.
N_Geneseq_36:T31255	+ 124.50	108.70	187.39	16812	HSV-2 strain S85 Contig ID 12
N_Geneseq_36:T31256	+ 124.00	125.90	20.63	1351	C. acidovorans gamma-lactamase
N_Geneseq_36:T31257	+ 124.00	124.32	25.27	2364	CD4-specific CDR-grafted heavy

N_Geneseq_36:T50962	+ 124.00	115.30	80.34	7073	TF8-569 CDR-grafted heavy c
N_Geneseq_36:T50963	+ 123.50	125.47	21.80	1929	Sorghum bicolor (L.) Moench
N_Geneseq_36:T50964	+ 123.00	125.54	21.60	1795	Human kidney injury related
N_Geneseq_36:T50965	+ 123.00	113.19	105.27	8051	Sugar biosynthesis gene clu
N_Geneseq_36:T50966	+ 123.00	113.08	106.78	8160	S. erythraea erythromycin-sy

seq_name: N_Geneseq_36:X28178

seq_documentation_block:

ID X28178 standard; cDNA; 2040 BP.
AC X28178;
DT 16-JUN-1999 (first entry)
DE Human PIGRL-1 coding sequence.
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;
KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;
KW diagnosis; therapy; ss.
OS Homo sapiens.
PN EP-905238-A2.
PD 31-MAR-1999.
PF 14-AUG-1998; 306487.
PR 30-OCT-1997; US-961564.
PR 25-AUG-1997; US-056935.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Sweet RW, Truneh A, Wu S;
DR WPI; 99-192666/17.
DR P-PSDB; Y05001.
PT New polypeptides encoding PIGRL-1 useful for treating diseases such
as X-linked Severe Combined Immunodeficiency
PS Claim 4; Page 18-19; 28pp; English.
CC This sequence encodes the human PIGRL-1 protein of the invention.
CC Autoimmune diseases involving altered expression or activity of PIGRL-1
may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined
Immunodeficiency (XSCID) and IgA deficiency. These diseases can be
diagnosed or susceptibility to them predicted by: (1) determining whether
there is a mutation in the genomic copy of the gene encoding PIGRL-1; or
(2) measuring the amount of PIGRL-1 in a sample derived from the patient.
Patients deficient in PIGRL-1 can be treated by administering either the
PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.
Patients with excessive expression or activity of PIGRL-1 can be treated
by administering an antagonist of PIGRL-1, an antisense nucleic acid
molecule which inhibits the expression of PIGRL-1 or administering
sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can
be used to identify its agonists by contacting a cell expressing PIGRL-1
with a candidate compound in the presence of a signal system and noting
the candidate as an agonist if a signal is produced. The same method can
be used to identify antagonists of PIGRL-1 but the presence of an
antagonist is indicated by a decrease in production of the signal.
Antibodies against PIGRL-1 may be used to isolate or identify clones
expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to
identify chromosomal mutations in the gene encoding PIGRL-1 in patients.
This information may then be correlated with the incidence of autoimmune
disease in those patients to identify whether the mutation causes the
disease.

alignment_scores:

Quality: 2047.00 Length: 390
Ratio: 5.262 Gaps: 0
Percent Similarity: 99.744 Percent Identity: 99.744

alignment_block:

US-09-135-238B-2 x X28178
Align seg 1/1 to: X28178 from: 1 to: 2040
1 MetaspArgTrpLeuTrpProLeuYrPheLeuProValSerGlyAlaLe 17
|||||
146 ATGGACCTCTGCGTTTGGCCACTTACTCTCCATGATCAGGGGCCCT 195
17 ATGGLLeuProGluValValGluGluLeuGlyGlySerValT 34
|||||
196 GAGGATCTCCCAAGAGTAAGGTAGAGGGGAGCTGGCGGATCAGTTA 245

34 hrileLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 50
 246 CCATCAAGTGCACCTTCCTGAAATGCATGTGAGGATATATCTGTGCGG 295
 51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67
 296 GAGATGGCTGGATCTGGACATGTGTACCTGGTATCCACACCAACTT 345
 67 eileLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgL 84
 346 CATCAAGGCAGAAATACAAGGGCGGAGTTACTCTGAAGCAATACCCACGCA 395
 84 ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSergly 100
 396 AGAATCTGTCTAGTGGAGTAACACAGCTGACAGAAAGTGACAGCGGA 445
 101 ValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnly 117
 446 GTCTATGCTCGGGAGCGGCATGAACACAGACCGGGGAAAGACCCAGAA 495
 117 sValThrLeuAsnValHisSerGluTyrGluProSerTrpGluGlnP 134
 496 AGTCACCCCTGAATGTCCACAGTGAATACGAGCCATCATGGGAAGACGAGC 545
 134 roMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMet 150
 546 CAATGCTGAGACTCCAAATGGTTTCACTGCGCTTATTTGTCCAGATG 595
 151 ProLAtyAlaSerSerSerLysPheValThrArgValThrProAl 167
 596 CTCTATGCTGAGACTTCCTCAAAATTCGTACACAGAGTTTACACACACGAGC 645
 167 aGlnArgGlyLysValProProValHisHisSerSerProThrThrGlnI 184
 646 TCAAAGGGGAGAGTCCCTCCAGTTCACCACTCTCTCCCGCCACCCCAAA 695
 184 leThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLys 200
 696 TCACCCACCGCCCTCGAGTGCCAGAGCATCTTCAGTAGCAGGTGACAAAG 745
 201 ProArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuG1 217
 746 CCCGGAACCTTCTGCCATCCACTACAGCCTCAAAATCTCAGCTCTGGA 795
 217 uGlyLeuLeuLysProGlnThrProSerTyrAsnHisThrArgLeuH 234
 796 GGGGCTCTCAGCCCCCAGAGCCGACGCTACACACACACACAGGCTGC 845
 234 isArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGln 250
 846 ACAGGCAGAGACACTGACTATGGCTCAGCTCTGGGAGGGAAGGCCAA 895
 251 GlyPheHisIleLeuIleProThrIleLeuGlyLeuPheLeuLeuAlaLe 267
 896 GGATTCACATCTCATCCCGACCATCTCTGGGCTTTCTGCTGGCACT 945
 267 uLeuGlyLeuValValLysArgAlaValGluArgArgGlyAlaLeuSerA 284
 946 TCTGGGGCTGGTGGTGAAGAGGCGCTTGAAGGAGGAAAGCCCTCTCCA 995
 284 rArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln 300
 996 GGGCGGCGCGCGACTGGCGGTGAGGATGCGCGCTTGAGAGCTCCACG 1045
 301 ArgProArgGlySerProArgProArgSerGlnAsnAsnIleTyrSerAl 317
 1046 AGGCCCCGGGGTCCCGCGACCGCGCTCCCAAAACACATCTACAGCGC 1095
 317 aCysProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlaP 334
 1096 CTGCCCCGCGCGCTCTGTGAGCGGACGCTGCAGGCACAGGGGAGGCC 1145

334 roValProGlyProGlyAlaProLeuProProAlaProLeuGlnValSer 350
 1146 CCGTCCCGCGCGGAGCGCGCTTGCCTCCCGCGCGCTGCAGGTGTCT 1195
 351 GluSerProTrpLeuHisAlaProSerLeuLysThrSerCysGluTyrVa 367
 1196 GAATCTCCCTGGCTCCATGCCCATCTCTGAGACACAGCTGTGAATAGT 1245
 367 lserLeuTyrHisGlnProAlaAlaMetMetGluAspSerAspSerAspA 384
 1246 GAGCTCTTACCACGAGCTCCCGCATGATGAGGAGCAGTGATTCAGATG 1295
 384 spTyrIleAsnValProAla 390
 1296 ACTACATCAATGTTCTCTGCC 1315
 seq_name: N_Geneseq_36:X28179
 seq_documentation_block:
 ID X28179 standard; cDNA; 1047 BP.
 AC X28179;
 DT 16-JUN-1999 (first entry)
 DE Human PIGRL-1 coding sequence fragment.
 KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;
 KW X-linked Severe Combined Immunodeficiency; XSCID; Iga deficiency;
 KW diagnosis; therapy; ss.
 OS Homo sapiens.
 PN EP-905238-A2.
 PD 31-MAR-1999.
 PF 14-AUG-1998; 306487.
 PR 30-OCT-1997; US-961564.
 PR 25-AUG-1997; US-056935.
 PA (SWIK) SMITHKLINE BEECHAM CORP.
 PI Sweet RW, Truneh A, Wu S;
 DR WPI: 99-192666/17.
 DR P-PSDB: Y05002.
 PT New polypeptides encoding PIGRL-1 useful for treating diseases such
 as X-linked Severe Combined Immunodeficiency
 PT Disclosure: Page 8; 26pp; English.
 CC This sequence encodes the human PIGRL-1 protein of the invention.
 CC Autoimmune diseases involving altered expression or activity of PIGRL-1
 CC may include hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined
 CC Immunodeficiency (XSCID) and Iga deficiency. These diseases can be
 CC diagnosed or susceptibility to them predicted by: (1) determining whether
 CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or
 CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.
 CC Patients deficient in PIGRL-1 can be treated by administering either the
 CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.
 CC Patients with excessive expression or activity of PIGRL-1 can be treated
 CC by administering an antagonist of PIGRL-1, an antisense nucleic acid
 CC molecule which inhibits the expression of PIGRL-1 or administering
 CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can
 CC be used to identify its agonists by contacting a cell expressing PIGRL-1
 CC with a candidate compound in the presence of a signal system and noting
 CC the candidate as an agonist if a signal is produced. The same method can
 CC be used to identify antagonists of PIGRL-1 but the presence of an
 CC antagonist is indicated by a decrease in production of the signal.
 CC Antibodies against PIGRL-1 may be used to isolate or identify clones
 CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to
 CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.
 CC This information may then be correlated with the incidence of autoimmune
 CC disease in those patients to identify whether the mutation causes the
 CC disease.
 SQ Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T;

alignment_scores:
 Quality: 1125.00 Length: 275
 Ratio: 4.429 Gaps: 12
 Percent Similarity: 92.364 Percent Identity: 90.182
 alignment_block:
 US-09-135-238B-2 x X28179

Align seg 1/1 to: X28179 from: 1 to: 1047

```
1 MetAspArgTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLeu 17
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
150 ATGAGCTCTGGCTTTGGCCACTTTTACTTCCGTGCCAGTATCAGGGCCCT 199
17 uArgIleLeuProGluValLysValGluGlyLeuGlyGlySerValT 34
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
200 GAGGATCTCCCAAGAGTAAGGTAGAGGGGAGCTGGCGGATCAGTTA 249
34 hrIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
250 CCATCAAGTGCCCACTTCTGAAATCATGTGAGGATATATCTGCGCG 299
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrAsnPh 67
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
300 GAGATGCTGGATCTGGACATGTGTGACCGTGGTATCCACCACCACTT 349
67 eIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgL 84
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
350 CATCAAGGCAGATAACAAGGCGGAGTTACTCTGAAGCAATACCCACGCA 399
84 ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
400 AGAATCTGTCTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGGA 449
101 ValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
450 GTCATGCTCGGGAGC.GGCATGAACACAGACCGGGGAAGACCCAGAA 498
117 sValThrLeuAsnValHisSerSerGlyTyrGluProSerTrpGluGlnP 134
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
499 AGTCACCTCGAATGTCCACAGTGAATACGAGCCATCATGGGAAGAGCAGC 548
134 rOMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMet 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
549 CAATGCTGAGACTCCAAATGGTTTCATCTGCTCCCTATTTGTTCCAGATG 598
151 ProAlaTyrAlaSerSerSerLysPheValThr..ArgValThrThrPro 166
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
599 CCTGATATCGCGGTCTTCCACATTCGTACCGCAGAGTTACCACACCA 648
167 Ala.GlnArgGlyLysValProValHisHisSerSerProThrThrG 183
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
649 GCTTCAAGGGGCAAGGTCCCTCCAGTTCACCACTCTCCCTCCGCCACCC 698
183 InIle..ThrHisArgPro.ArgValSerArgAlaSerSerValAlaGlyA 199
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
699 AAATTCACCCACCGCCCTTCGAGTGTNCAGAGCATCTTCAGTAGCAGGTG 748
199 sPLysProArg..ThrPheLeuProSerThrThrAlaSerLysIleSerAl 215
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
749 ACAAGCCCGCAAACTTCTCGCATCCACTACAGCCTCAAAATCTCAGC 798
215 aLeuGluGlyLeuLeuLys....ProGluThrProSerTyrAsnHisHist 231
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
799 YCTGAAGGGGTGCTTCAAGCCCGCCAGAGCGCCAGCTACAAACAN.CACA 847
231 hrArgLeuHisArgGlnArgAlaLeuAsp...Tyr.GlySerGlnSerGl 246
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
848 CCAGGCTGCACAGGACAGACACTGGATATCTTATGGNTCCACAGTCTGG 897
246 yArgGluGly.....GlnGlyPhe...HisIleLeuIlePro..ThrIle 259
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
898 G...GAGGGGAANGNCCAAGGATTTNACATTCCTGATTCCTCCGACCATC 944
260 LeuGlyLeuPhe 263
||||| |||
945 NVTGGGCGCTTT 956
```

seq_name: N_Geneseq_36:T31290

```
seq_documentation_block:
ID T31290 standard; CDNA; 3095 BP.
AC T31290;
DT 24-FEB-1997 (first entry)
DE Mouse poly-immunoglobulin receptor, cDNA.
KW Mouse; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; murine; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT 85..2400
FT cds /*tag= a
FT
PN W09621012-Al.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNKE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI; 96-333987/33.
DR P-PSDB; W03180.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure: Pages 117-121; 152pp; English.
CC The present sequence encodes the mouse poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-320, 450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a pp as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The pp protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 3095 BP; 861 A; 796 C; 784 G; 654 T;
```

alignment_scores:

Quality:	198.00	Length:	168
Ratio:	1.980	Gaps:	5
Percent Similarity:	59.524	Percent Identity:	31.548

alignment_block:

US-09-135-238B-2 x T31290 ..

Align seg 1/1 to: T31290 from: 1 to: 3095

```
10 PheLeuProValSerGlyAlaLeuArgIleLeuProGluValLysValGl 26
||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
121 TTTTCAGGGTCTCCACAAAAGCCCATTTGGTCCCCCAGGAGTGAG 170
26 uGlyGluLeuGlyGlySerValThrIleLysCysProLeuProGlu.... 41
: : : ||| ||||||| ||| ||| |||
171 TAGTATAGAGGCGACTCTGTTTCCATCACGTGCTACTACCCAGACACCT 220
42 .....MetHisValArgIleTyrLeuCysArgGluMetAlaGlySer 55
||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
221 CTTCTAACCGGCACACCCGGAATACTGTCGCCGACAAGGAGCC...AGC 267
56 GlyThrCysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyr 72
||| ||| ||| : : : ||| : : : ||| : : : ||| : : : |||
268 GGCATGTGCACAACGCTCATCTCTCAATGGCTACCTCTCCAAGAGTA 317
72 rLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuV 89
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
```

```

318 TTCAGGCGAGCCAACTCATCACTCCAGAGAACACACATTTGTGA 367
89  algluValThrGlnLeuThrGluSerAspSerGlyValTyAlaCysGly 105
      ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
368  TTACATTGAGCAGCTCACCAGGACGACACTGGGAGCTACAACTGTGGC 417
      |||  ::::: ||||| ::::: ||||| ::::: |||||
106  AlaglyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVa 122
      |||  ::::: ||||| ::::: ||||| ::::: |||||
418  CTGGGT...ACCAGTAACCGAGGCTGTCTTCGATGTCAGCTGGAGGT 464
      |||  ::::: ||||| ::::: ||||| ::::: |||||
122  lHisSerGluTyGluProSerTrpGluGluGlnProMetProGluThrP 139
      |  ::::: ||||| ::::: ||||| ::::: |||||
465  C.....AGCCAGGTCTCTGAGTTGC 484
      |  ::::: ||||| ::::: ||||| ::::: |||||
139  rOLysTrpPheHisLeuProTyLeuPheGlnMetProAlaTyAlaSer 155
      ||::: ||||: ||||: ||||: ||||: ||||:
485  CGAGTGACCCACGTC.....TACACAAAG 510
      ||::: ||||: ||||: ||||: ||||: ||||:
156  SerSerLysPheValThrArgValThrThrProAlaGlnArgGlyLysVa 172
      :::  :::  |||  ::::: ||||  ::::: ||||
511  GACATAGCGAGAAATGTGACATTGAATGCCCTTTCAAAGGAGGAGATGT 560
      |||  ::::: ||||| ::::: ||||| ::::: |||||
172  lPro 173
      ||||
561  TCCC 564

seq_name: N_Geneseq_36:V15422
seq_documentation_block:
ID  V15422 standard; DNA; 29392 BP.
AC  V15422;
DT  11-JUN-1998 (first entry)
DE  Mouse poly Ig receptor protein gene.
KW  Mouse; poly Ig receptor protein; pigr protein; pig; deficiency;
KW  knockout mouse; disease model; ds.
OS  Mus sp.
PN  J10057066-A.
PD  03-MAR-1998.
PF  19-AUG-1996; 217154.
PR  19-AUG-1996; JP-217154.
PA  (HONS ) YAKULT HONSHA KK.
DR  WPI; 98-254323/23.
PT  Mouse pig receptor protein gene - used for preparing gene knockout
PT  mice, useful for study of human poly Ig receptor protein deficiency
PS  Claim 1; Page 4-14; 18pp; Japanese.
CC  The present sequence represents the mouse poly Ig receptor protein
CC  gene, which has a 29392 bp sequence. The new gene can be used to
CC  produce a gene knockout mouse, useful as a disease model of human
CC  poly Ig receptor protein deficiency.
SQ  Sequence 29392 BP; 8318 A; 6747 C; 6514 G; 7813 T;

alignment_scores:
  Quality: 188.00      Length: 97
  Ratio: 2.648        Gaps: 3
Percent Similarity: 73.196 Percent Identity: 40.206

alignment_block:
US-09-135-238B-2 x V15422 ..
Align seg 1/1 to: V15422 from: 1 to: 29392

30  GlySerValThrIleLysCysProLeuProGlu.....Me 42
    ||| ||||| ::::: |||||
18245  GCGGACTCTGTTCCACTGCTACTACCCAGACACACTCTGTCAACG 18294
      ||| ||||| ::::: |||||
42  tHisValArgIleTyLeuCysArgGluMetAlaGlySerGlyThrCysG 59
    ||||| ::::: ||||| ::::: ||||| ::::: |||||
18295  GCACACCCGGAATCTGGTCCGACAGGAGGCC...AGCGGATGTGCA 18341
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
59  lyThrValValSerThrThrAsnPheIleLysAlaGluTyLysGlyArg 75
    ||||| ::::: ||||| ::::: ||||| ::::: |||||
18342  CAAGGCTCATCTCTCAAAATGGTACCTCTCCAAAGGAGTATTCAGGCAGA 18391

```

```

76  ValThrLeuLysGlnTyProArgLysAsnLeuPheLeuValGluValTh 92
      ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
18392  GCCAACTCATCACTCCAGAGAACACACATTGTGATTAACTTGA 18441
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
92  rGlnLeuThrGluSerAspSerGlyValTyAlaCysGlyAlaGlyMeta 109
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
18442  GCAGCTCACCCAGGACGACCTGGGAGCTACAACTGTGGCTGGGT...A 18488
      ::::: ||||| ::::: ||||| ::::: |||||
109  snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
      ::::: ||||| ::::: ||||| ::::: |||||
18489  CCAGTAACCGAGGCTGTCTTCGATGTCAGCTGGAGGTC 18529
      ::::: ||||| ::::: ||||| ::::: |||||

seq_name: N_Geneseq_36:V20383
seq_documentation_block:
ID  V20383 standard; cDNA to mRNA; 1911 BP.
AC  V20383;
DT  26-JUN-1998 (first entry)
DE  cDNA for human immunity related factor.
KW  Lymph node; human; immunity related factor; research; treatment;
KW  immune disease; infectious disease; ds.
OS  Homo sapiens.
PN  Location/Qualifiers
FT  CDS
FT  sig_peptide
FT  mat_peptide
FT  product= immunity_related_factor

J10072495-A.
PD  17-MAR-1998.
PF  11-JUN-1997; 153218.
PR  13-JUN-1996; JP-152362.
PA  (ASAH ) ASAH KASEI KOGYO KK.
DR  WPI; 98-234766/21.
DR  P-PSDB; W50033.
PT  Immunity related factor - useful in the treatment of immune related
PT  and infectious diseases
PS  Claim 15; Pages 18-20; 21pp; Japanese.
CC  The present sequence encodes a lymph node derived human immunity
CC  related factor, which can be used to research and treat immune and
CC  infectious diseases.
SQ  Sequence 1911 BP; 490 A; 541 C; 525 G; 355 T;

alignment_scores:
  Quality: 187.50      Length: 575
  Ratio: 0.906        Gaps: 20
Percent Similarity: 36.000 Percent Identity: 18.957

alignment_block:
US-09-135-238B-2 x V20383 ..
Align seg 1/1 to: V20383 from: 1 to: 1911

3  ArgTrpLeuTrpProLeuTyPheLeuProValSerGlyAlaLeuArgI1 19
    ||||| ::::: ||||| ::::: |||||
170  AGATGGCTGTGGAG...GGCTCTCTCCCTCCAGGACCCATCTCCGGGC 216
      ||||| ::::: ||||| ::::: |||||
19  eLeuProGluValLys..... 24
    ::::: ||||| ::::: |||||
217  CATGGGAACACTCAGGCTCTCTCGCCCTCTCTGCGGGGAGGAGAGCT 266
      ||||| ::::: ||||| ::::: |||||
25  .....ValgluGlyGlu 28
    ||||| ::::: ||||| ::::: |||||
267  CTTTTCAGCTCCAAATTCATTGAAGGGCTCAAGGCTGTGTCTCAGGGAG 316
      ||||| ::::: ||||| ::::: |||||
29  LeuGlyGlySerValThrIleLysCysProLeuPro.....G1 41
    ||||| ::::: ||||| ::::: |||||
317  CTGGAGGAGCTGTCCACATCCAGTGCCTATATGCCCTCATCTGTCTCAA 366

```



```

41 uMethisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrC 58
   :   |||   |||   |||   |||   |||   |||   |||   |||
367 CAGGCACCAAGGAAGTAGTACTGTCGCGTCTGGGGCCCCCAAGATGGATCT 416
   :   |||   |||   |||   |||   |||   |||   |||   |||
58  yGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGly 74
   ||   |||   |||   |||   |||   |||   |||   |||   |||
417 GCCACCAATTGTGTCCACCAACCACTATATCTACCATCGCTATCGTGAC 466
   ||   |||   |||   |||   |||   |||   |||   |||   |||
75  ArgValThrLysGlnTyrProArgLysAsnLeuPheLeuValGluVa 91
   ||   |||   |||   |||   |||   |||   |||   |||   |||
467 CGTGTGGCCCTCAGAGATTCCACAGAGAGCITGTGTGGTGAGGCT 516
   ||   |||   |||   |||   |||   |||   |||   |||   |||
91  lThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyM 108
   ||   |||   |||   |||   |||   |||   |||   |||   |||
517 GTCCCAACTGTCCCGGATGACATCGGATGCTACCTCTGCGGCATTGAA 566
   ||   |||   |||   |||   |||   |||   |||   |||   |||
108 eAsnThrAspArg..... 112
   ||   |||   |||   |||   |||   |||   |||   |||   |||
567 GTGAACAACAATGCTGTTCCTTAAGCATGAATCTGACCATCTCGAGGT 616
   ||   |||   |||   |||   |||   |||   |||   |||   |||
112 ..... 112
617 CCGGCCAGCACCCCTCCACAGCCACTCCAGCTGCTGGGAGCTCACCAT 666
   ||   |||   |||   |||   |||   |||   |||   |||   |||
113 .....GlyL 114
   ||   |||   |||   |||   |||   |||   |||   |||   |||
667 GAGATPCTTGAACAGCGCTCTCCAGTGGCCACAGATGGACCCACGAA 716
   ||   |||   |||   |||   |||   |||   |||   |||   |||
114 ysThrGln..... 116
   ||   |||   |||   |||   |||   |||   |||   |||   |||
717 CCACCCAGACTTAGCAGGGGACAGCATGGGACACAGTTGCTTCCACT 766
   ||   |||   |||   |||   |||   |||   |||   |||   |||
117 .....LysValThrLeuAsnValHisSerGluTyrGluPro.. 128
   ||   |||   |||   |||   |||   |||   |||   |||   |||
767 CCAGGAACCAAGACAGACTACAGCTTCAGCTCAGGGAAGACCAACCCAGG 816
   ||   |||   |||   |||   |||   |||   |||   |||   |||
129 .....SerTrpGlu..... 132
   ||   |||   |||   |||   |||   |||   |||   |||   |||
817 AGCAACAGCGCCAGCAGCTCCAGGAGACAGGCGCTGGCGCAGGGTCTG 866
   ||   |||   |||   |||   |||   |||   |||   |||   |||
133 .....GlnProMetProGluThrProLysTrpPheHisLeuPro 145
   ||   |||   |||   |||   |||   |||   |||   |||   |||
867 TCANAGCACTGCTCCGATCCAGAGATCCA..... 898
   ||   |||   |||   |||   |||   |||   |||   |||   |||
146 TyrLeuPheGlnMetProAlaTyrAlaSerSerLysPheValThrAr 162
   ||   |||   |||   |||   |||   |||   |||   |||   |||
899 .....CCTTCAAGAGCAGAGCATGTCCAA 924
   ||   |||   |||   |||   |||   |||   |||   |||   |||
162 gValThrThrProAlaGlnArgGlyLysValProProValHisHisSerS 179
   ||   |||   |||   |||   |||   |||   |||   |||   |||
925 TACAACAGAAGGTGTTTGGGAGGC..... 949
   ||   |||   |||   |||   |||   |||   |||   |||   |||
179 erProThrThrGlnIleThrHisArgProArgValSerArg..... 192
   ||   |||   |||   |||   |||   |||   |||   |||   |||
950 ..ACCAGAGCTCGGTGACAAACAGGCTAGAGCCACCAAGGACGAGGAGG 997
   ||   |||   |||   |||   |||   |||   |||   |||   |||
193 ...AlaSerSerValAlaGlyAspLysProArg..... 202
   ||   |||   |||   |||   |||   |||   |||   |||   |||
998 GAGATGACAACTACCAAGGCTGATAGGCCAAGGGAGGACATAGAGGGGT 1047
   ||   |||   |||   |||   |||   |||   |||   |||   |||
202 ..... 202
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1048 CAGGATAGCTCTTGATGACGCCAAAAAGTCTTAGAACCATTTGGGCCAC 1097
   ||   |||   |||   |||   |||   |||   |||   |||   |||
203 .....ThrPheLeuProSerThrThr 209
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1098 CAGCTCTGGTCTCAGAACCTTGGCCTGGGAATCTCCACAGCAACG 1147
   ||   |||   |||   |||   |||   |||   |||   |||   |||
210 AlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProGlnThrProSe 226
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1148 CCAGTTTCTAAGCAACAATCTCAGGGTTCCATTGGAGAACAACTCCAGC 1197
   ||   |||   |||   |||   |||   |||   |||   |||   |||
226 rTyrAsnHisHisThrArg..... 232

```

```

1198 TGCAGGATGTGGACCTTGGGAACCTCCAGCTGCAGATGTGTGGATCTTGG 1247
   :   |||   |||   |||   |||   |||   |||   |||   |||
233 .....LeuHisArgGlnArgAlaLeuAspTyr 241
   :   |||   |||   |||   |||   |||   |||   |||   |||
1248 GAACCTCCAGCTGCAGATGTGTGACCAGCATGGAGGCACATCTGGGGAA 1297
   :   |||   |||   |||   |||   |||   |||   |||   |||
242 GlySerGlnSerGly..... 246
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1298 GGAAGCGCTCGAGGGACCTAGATGCTGCCACTGGAGACAGAGGTCCCCA 1347
   :   |||   |||   |||   |||   |||   |||   |||   |||
246 ..... 246
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1348 AGCAACACTGAGCCACACCCCGCAGTAGACCTGGGACCCCTGGCA 1397
   :   |||   |||   |||   |||   |||   |||   |||   |||
247 .....ArgGluGlyGlnGlyPheHis 253
   :   |||   |||   |||   |||   |||   |||   |||   |||
1398 AGGAGTCTCCGTGAAGCGTACTTTTCCAGAAAGATGAAAGCAGCTCGG 1447
   :   |||   |||   |||   |||   |||   |||   |||   |||
254 IleLeuIlePro.....ThrIleLeuGlyLeuPheLeuAlaLeuLe 268
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1448 ACCTGCTGCTCTGTCTCTACCATGCTGGCCCTGTTTATGCTTATGCTCT 1497
   :   |||   |||   |||   |||   |||   |||   |||   |||
268 uGlyLeuValValLysArgAlaValGluArgArgLysAlaLeuSerArgA 285
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1498 GGTCTCTATTG.....CAAAGGAAGCTCTCGAGAAGA 1529
   :   |||   |||   |||   |||   |||   |||   |||   |||
285 rgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln... 300
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1530 GGACCTCTCAGGAGGCAGAAAGGTCACCTTAATTCAGATGACACATTTT 1579
   :   |||   |||   |||   |||   |||   |||   |||   |||
301 .....ArgProArgGlySerProArgProArgSerGlnAsnAl 314
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1580 CTGGAAGTGAACCCCAAGCAGACCCAGCTGCCCATGTGGAAGAAGAAT 1629
   :   |||   |||   |||   |||   |||   |||   |||   |||
314 eTyr.....SerAlaCysProArgArgAlaArgGlyAlaAspAlaAlaG 329
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1630 GCTCCAGGATGACTCTCTCTCT.....GCTGGGGCCAGGC 1664
   :   |||   |||   |||   |||   |||   |||   |||   |||
329 lyThrGlyGluAlaProValProGlyPro.....Gly 339
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1665 TGACTGCCCCAGAGAGAAATCCAGGACCTGAGGAGACAGAGATGAACT 1714
   :   |||   |||   |||   |||   |||   |||   |||   |||
340 AlaProLeuPro.....ProAl 345
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1715 GTCAGTTACCATGGAGAAGCAACCAAGATCAAGGCCTTCAGGACCCCA 1764
   :   |||   |||   |||   |||   |||   |||   |||   |||
345 aProLeuGlnValSerGluSerPro 353
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1765 GCCTCTTTCATCATCTCTCTCTCCA 1789
   :   |||   |||   |||   |||   |||   |||   |||   |||

seq_name: N_Geneseq_36:T31291
seq_documentation_block:
ID   T31291 standard; cDNA; 3269 BP.
AC   T31291;
DT   24-FEB-1997 (first entry)
KW   Rat poly-immunoglobulin receptor, cDNA.
DE   heavy chain; immunoglobulin receptor; protection protein; mutants;
KW   mucosal; antigen binding domain; protection; pathogen;
KW   Guy's 13 antibody; gastrointestinal; passive; immunisation;
KW   poly; sorbinus; ss.
OS   Rattus rattus.
FH   Key
FT   cds
FT   Location/Qualifiers
     74..2383
        /*tag= a
WO9621012-A1.
PN   11-JUL-1996.
PD   27-DEC-1995; U16889.
PF   30-DEC-1994; US-367395.
PR   04-MAY-1995; US-434000.
PA   (PLAN-) PLANT BIOTECHNOLOGY INC.

```

PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.

PA (PLAN-) PLANET BIOTECHNOLOGY INC.

PI Hiatt AC, Lehner T, Ma JKC;

DR WPI: 96-333987/33.

DR P-PSDB: W03181.

PT Immunoglobulin and protection protein complex and its prodn. in plants - useful for passive immunisation against mucosal antigens, esp. against S. mutans and S. sorbinus to prevent dental caries (Disclosure; Pages 123-127; 152pp; English.

PS The present sequence encodes the rat poly-immunoglobulin (Ig)

CC receptor, a portion of which corresp. to residues 1-627, pref. 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,

CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).

CC The Ig of the invention comprises a PP as above in association with an Ig derived heavy chain, having at least a portion of an antigen

CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.

CC gastrointestinal, environments, therefore enhancing its

CC effectiveness in passively immunising animals against mucosal pathogens. The Ag binding domain is specifically derived from the

CC Guy's 13 antibody, and the Ig can be used to prevent dental caries

CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or

CC S. sorbinus serotypes d and g.

SQ Sequence 3269 BP; 890 A; 831 C; 850 G; 698 T;

alignment_scores:

Quality: 185.50 Length: 164

Ratio: 1.855 Gaps: 6

Percent Similarity: 60.976 Percent Identity: 30.488

alignment_block:

US-09-135-238B-2 x T31291 ..

Align seg 1/1 to: T31291 from: 1 to: 3269

30 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 42

170 GGTACTCGGTCTCCATCAGTGTCTACTACCAGACACCTCTGTCAACCG 219

42 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59

220 GCACACCCGGAAATACTGTGTCGACAGAGGCC...AACGGCTACTGCG 266

59 lyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 75

267 CAACCTCATCTCTCAATGCTTACCTCTCGAAGGAGTATTCAGGCAGA 316

76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92

317 GCCAGCCTCATCAACTCCAGAGAAATAGCACATTTGTGATTAACATTGC 366

92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 109

367 ACATCTCACCCAGGAGGACACTGGGAGCTACAGTGTGCTGGGT...A 413

109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125

414 CCACCTAACCGAGGCGCTGTTTTCGATGTCAGCGCTGGAGGTC..... 454

126 TyrGluProSerTyrGluGlnProMetProGluThrProLysTyrPpH 142

455AGCCAGGTTCTCTGAGTTCCTCCAAATGACAC 483

142 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerLysP 159

484 CCATGTC.....TACACAAAGG 500

159 heValThrArgValThrProAlaGlnArgGlyLysValProVal 175

501 ACATAGGACAGAACTGTGACCATCGAATGCCGTTTCAAGAGGGGAATGCT 550

176 HisHisSer.SerProThrThrGlnIleThrHisArgPro 188

||| ...:||||| :|:|:|

551 CATAGCAGAAATCCCTGTGTAAGAGAGAGAGAGGCCT 590

seq_name: N_Geneseq_36:T30857

seq_documentation_block:

ID T30857 standard; cDNA: 1936 BP.

AC T30857:

DT 13-SEP-1996 (first entry)

DE Secretary component DNA fragment 2.

KW PolymERIC immunoglobulin SC DNA fragment 2; crystallisation;

OS Chimeric Homo sapiens;

OS Chimeric synthetic.

PH Key Location/Qualifiers

FT misc_feature 1..46

FT /tag= a /note= "bases 1-46 derived from PCR with 5'

FT primer 2"

FT mat_peptide 16..1920

FT /tag= b

FT /product= transcript from SC DNA fragment 2

FT misc_feature complement (1902..1936)

FT /tag= c

FT /note= "from PCR with 3' primer"

FN WO9618734-A1.

PD 20-JUN-1996.

PF 06-DEC-1995; E04797.

PR 16-DEC-1994; EP-120019.

PA (CIBA) CIBA GEIGY AG.

PI Asselbergs FAM, Gruetter M, Leist C, Van Oostrum J;

DR WPI: 96-300651/30.

PT Prodn. of recombinant secretory protein capable of crystallisation -

PT Pref. human poly-IgR, for use in screening and binding studies

PS Claim 5; Page 39-41; 50pp; English.

CC SC DNA fragment 2 (T30857) was obtd. by PCR amplification of human

CC polymERIC immunoglobulin receptor (pIgR) cDNA in vector pCB6 using a

CC 5' primer (T30860) that creates an HindIII site upstream of the AAG

CC start codon and a 3' primer (T30861) that generates a stop codon at

CC position 1906 just upstream of the hydrophobic transmembrane segment

CC Of the pIgR and also creates a downstream XbaI site. CHO SSF 3

CC cells transfected with a vector carrying the amplified DNA can be

CC used for prodn. of recombinant secretory component (SC). This can

CC be crystallised for receptor structure studies, used to stabilise

CC immunoglobulins, or used to screen (ant)agonists capable of

CC modulating mucosal immune responses.

SQ Sequence 1936 BP; 452 A; 525 C; 578 G; 381 T;

alignment_scores:

Quality: 185.00 Length: 458

Ratio: 0.944 Gaps: 20

Percent Similarity: 42.795 Percent Identity: 22.489

alignment_block:

US-09-135-238B-2 x T30857 ..

Align seg 1/1 to: T30857 from: 1 to: 1936

12 ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGly 28

76 CCCATTATTGTT.....CCCGAG...GAGGTGAATAGTGT 107

28 uLeuGlyGlySerValThrIleLysCysProLeuPro.....G 41

108 GGAAGGTAACTAGTCTCATCGTGTACTACCCACCCACCTCTGTCTCA 157

41 luMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThr 57

158 ACCGGCACACCCGGAAGTACTGTGTCGCGCAGGAGCTAGAGTGGC... 204

58 CysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysG 74

205 TGCATAACCCCTCATCTCTCGGAGGCTACGTCTCCAGCAATATGCAGG 254

```
74 yArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluV 91
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
255 CAGGCTACCTCACCACCTCCCGGAGACGGCACATTCGTGGTGACCA 304
91 alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
305 TTGCCAGCTGAGCCAGGATGACTCCGGCGCTCAAGTGTGGCTGGC 354
108 MetAsnThr..... 110
:|||||:
355 ATCAATAGCCGAGGCGCTGCTTTGATGTCAGCCTGGAGGTCAGCCAGG 404
111 .....AspArgGlyLysT 115
|||||:|||||:
405 TCCTGGCTCCTAATGACACTAAAGTCTACACAGTGGACCTGGCGAGA 454
115 hrGlnLysValThrLeuAsnValHisSerGluTyrGluProSerTrpGlu 131
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
455 CG.....GTGACCATCAACTGCCCTTTCAAGACTGAGATGCTCAAAAG 498
132 GluGln.ProMetProGlu.....ThrProLysTrpPheHisL 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
499 AGGAAGTCTTGTACAAGCAGATAGGCGTGTACCCCTGTGCTGTCATCGA 548
144 euProTyrLeuPheGlnMetProAlaTyrAlaSerSerLysPheVal 160
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
549 CTCACGTGTTATGTGAATCCCACTATACAGGAAGATAGCCCTTGATA 598
161 ThrArgVal..... 163
|||||:
599 TTCAGGCTACTGCCAGTACTGTTTCAGCGTGTGTCATCAACCAACTCAGG 648
164 .....ThrThrProAlaGlnArgGlyLysValProProv 175
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
649 CTCACGATGCTGGCGATATCTCTGCCAGGCTGGGGATGATTCCAATAG 698
175 alHisHisSerSerProThrThrGlnIleThrHisArgProArgValSer 191
|||:|||||:
699 TAATAAGAAGAATGCTGACC..... 718
192 ArgAlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerTh 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
719 .....TCCAAGTCTAAAGCCGACCGCGAGCTGGTTATGAAGACCTC 762
208 rThrAlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProGlnThr. 224
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
763 AGGCGCTCAGTGACCTCCACTGTGCCCTGGCCCTGAGGTGGCAACGT 812
225 ..ProSerTyrAsnHisHisThrArgLeuHisArg..... 235
|||||:|||||:
813 GGCCAAATTTCTGTCCCGACAGACAGTCGTGGGAAACCTGTGCGTGC 862
236 .....GlnArgAlaLeuAspTyrGlySerGlnSe 245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
863 TCAACACCCCTAGGAGAGGCGCCCGCTTTGAGGCGCAGGATCCTGCTC 912
245 r.....G 246
|
913 AACCCCGAGGACAGGATGGCTCATTCAGTGTGGTGATCACAGGCTGAG 962
246 lyArgGluGlyGlnGlyPheHisIleLeuLeuProThrIleLeuGlyLeu 262
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
963 GAAGGAGGATGCAGGCGCTACCTGTGTGGAGCCCAATT..... 1000
263 PheLeuLeuAlaLeuLeuGlyLeuValValLysArgAlaValGluArgAr 279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1001 .....CGGATGTCAGCTGCAG 1017
279 gLysAlaLeuSerArgArgAlaArgLeuAlaValArgMetArgAlaL 296
|||||:|||||:
1018 GAAGGCT.....CGCCTATCCAGGC 1037
```

```
296 euGluSerSerGlnArgProArgGlySerProArgProArgSer..... 310
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1038 CTGGCAACTCTTCGTCATGAGAGTCCACCATTCCTCCCGCAGCCCATCG 1087
311 .....GlnAsnAsnIleTyr...SerAlaCysProArgAr 321
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1088 TGGTGAAGGGGTGGCAGGAGCTCTGTGGCGTCTGCTGCCCTACAAAC 1137
321 gAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlaProValProGlyP 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1138 CGTAAGAAAGCAAAAGCATCAAGTACTGGT.....GTCTCTGGGA 1178
338 roGlyAlaProLeuProAlaPro..... 346
|||||:|||||:
1179 AGGGCCCAAGATGGCGCTGCCCTCTGCTGGTGGACAGCGAGGGTGG 1228
347 .....LeuGlnValSerGluSerProTyr.....LeuHi 356
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1229 TTAAGGCCCACTAGCAGGCGCGCTCTCCCTGCTGGAGGAGCCAGGCAC 1278
356 salaProSerLeuLysThrSer 363
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1279 GGCACCTTCACCTGTCATCTCA 1300
seq_name: N_Geneseq_36:X07407
seq_documentation_block:
ID X07407 standard; cdna to mRNA; 1839 BP.
AC X07407;
DT 08-JUN-1999 (first entry)
DE Human secretory Immunoglobulin A component gene.
KW Immunoglobulin A; secretory; component; IGA; human; treatment;
KW prevention; infection; HIV; AIDS; cold; flu; virus; gene;
KW human immunodeficiency virus; respiratory syncytial virus; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..1839
FT /tag= a
FT /product= secretory component
FN WO9857993-A1.
PD 23-DEC-1998.
PF 10-JUN-1998; U11975.
PR 19-JUN-1997; US-050969.
PA (REGC ) UNIV CALIFORNIA.
PI Chintalacharuvu KR, Morrison SL;
DR WPI; 99-080950/07.
DR P-PSDB; W95601.
PT Producing secretory immunoglobulin in single cells - useful to
PT produce commercial quantities of secretory immunoglobulin to prevent
PT or treat infections
PS Disclosure; Page 22; 39pp; English.
CC The sequence is that of the coding region for the secretory
CC component of human secretory immunoglobulin A (siga).
CC The sequence can be used as part of a method for the
CC production of sig molecules. This method is useful for
CC producing commercial quantities of sig (especially siga) to treat
CC or prevent infections. In particular, siga produced by the method
CC can be used to prevent or treat infections in mammals, birds or
CC fish; especially systemic infections or infections at a mucosal
CC surface. It is especially useful to prevent or treat infection
CC with human immunodeficiency virus (HIV), respiratory syncytial
CC virus, flu virus or cold virus. The method allows production of
CC commercial quantities of sig molecules for therapeutic use, not
CC previously possible; production using non-plant cells and a
CC single cell type is more efficient than a previous multi-step
CC process of fusing recombinant plant cells, and avoids alterations
CC of the sig by plant cells. SIGA molecules are more stable
CC and resistant to proteolysis than previously used IgA molecules,
CC and can be administered to prevent as well as to treat infections,
CC unlike e.g. IgG and IgM molecules.
CC Sequence 1839 BP; 423 A; 504 C; 544 G; 368 T;
SQ
```

```

alignment_scores:
  Quality: 182.00      Length: 115
  Ratio: 2.247        Gaps: 5
  Percent Similarity: 70.435  Percent Identity: 37.391

alignment_block:
  US-09-135-238B-2 x X07407  ..

Align seg 1/1 to: X07407 from: 1 to: 1839

12 ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGlyL 28
   |||||
61 CCCAATTTGGT.....CCCGAG...GAGGTGAATAGTGT 92
28 uLeuGlySerValThrIleLysCysProLeuPro.....G 41
   |||||
93 GGAGGTAACCTCAGTGCATCAGTGCCTACTACCCACCCACCTCTGTCA 142
41 luMetHisValArgIleLeuCysArgGluMetAlaGlySerGlyThr 57
   |||||
143 ACCGGCACCCCGAAGTACTGGTCCCGGAGGAGCTAGAGTGGC... 189 a
58 CysGlyThrValSerThrThrAsnPheIleLysAlaGluTyrLysG 74
   |||||
190 TGCATAACCTCATCTCCTCGGAGGCTAGTCTCCAGCAAAATATGCAG 239
74 yArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValG 91
   |||||
240 CAGGCTAACCTCACCACCTCCCGGAGGAGGCACATTTGTGTGAACA 289
91 alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaG 107
   |||||
290 TTGCCACAGCTCAGCAGGATGACTCCGGGCGCTACAAGTGTGGCTGG 339
108 MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
   |||||
340 ATCAATAGC...CGAGGCTGTCTTTCATGTCAGGCTGGAGGTC 381

seq_name: N_Geneseq_36:T30856

seq_documentation_block:
ID T30856 standard; cDNA; 2031 BP.
AC T30856;
DT 13-SEP-1996 (first entry)
DE Secretory component DNA fragment 1.
KW Secretory component; SC DNA fragment 1; crystallisation;
OS Polymeric immunoglobulin receptor; pigR; CHO; ss.
OS Chimeric Homo sapiens;
OS Chimeric synthetic.
FH Key Location/Qualifiers
FT misc_feature 1..33
FT /tag= a
FT /note= "bases 1-33 derived from PCR with 5'
FT primer 1"
FT lll..2015
FT mat_peptide /tag= b
FT /product= transcript from SC DNA fragment 1
FT misc_feature complement (1997..2031)
FT /tag= c
FT /note= "from PCR with 3' primer"
FN W09618734-A1.
PD 20-JUN-1996.
PF 06-DEC-1995; E04797.
PR 16-DEC-1994; EP-120019.
PA (CIBA ) CIBA GEIGY AG.
PI Asselbergs FAM, Gruetter M, Leist C, Van Oostrum J;
PR WPI: 96-300651/30.
DT Prodn. of recombinant secretory protein capable of crystallisation -
PT pref. human poly-IgR, for use in screening and binding studies
PS Claim 5; Page 36-38; 50pp; English.
CC SC DNA fragment 1 (T30856) was obt'd. by PCR amplification of human
CC polymeric immunoglobulin receptor (pigR) cDNA in vector pcB6 using a
CC 5' primer (T30859) hybridising to the cytomegalovirus promoter in

```

```

CC pcB6 and a 3' primer (T30861) that generates a stop codon at
CC position 1906 just upstream of the hydrophobic transmembrane segment
CC of the pigR and also creates a downstream XbaI site. CHO SSF 3
CC cells transfected with a vector carrying the amplified DNA can be
CC used for prodn. of recombinant secretory component (SC). This can
CC be crystallised for receptor structure studies, used to stabilise
CC immunoglobulins, or used to screen (ant)agonists capable of
CC modulating mucosal immune responses.
SQ Sequence 2031 BP; 477 A; 541 C; 605 G; 408 T;

```

```

alignment_scores:
  Quality: 182.00      Length: 115
  Ratio: 2.247        Gaps: 5
  Percent Similarity: 70.435  Percent Identity: 37.391

```

```

alignment_block:
  US-09-135-238B-2 x T30856  ..

Align seg 1/1 to: T30856 from: 1 to: 2031

12 ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGlyL 28
   |||||
171 CCCAATTTGGT.....CCCGAG...GAGGTGAATAGTGT 202
28 uLeuGlySerValThrIleLysCysProLeuPro.....G 41
   |||||
203 GGAAGTAACCTCAGTGCATCAGTGCCTACTACCCACCCACCTCTGTCA 252
41 luMetHisValArgIleLeuCysArgGluMetAlaGlySerGlyThr 57
   |||||
253 ACCGGCACCCCGAAGTACTGGTCCCGGAGGAGCTAGAGTGGC... 299
58 CysGlyThrValSerThrThrAsnPheIleLysAlaGluTyrLysG 74
   |||||
300 TGCATAACCTCATCTCCTCGGAGGCTAGTCTCCAGCAAAATATGCAG 349
74 yArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValG 91
   |||||
350 CAGGCTAACCTCACCACCTCCCGGAGGAGGCACATTTGTGTGAACA 399
91 alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107
   |||||
400 TTGCCACAGCTCAGCAGGATGACTCCGGGCGCTACAAGTGTGGCTGG 449
108 MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
   |||||
450 ATCAATAGC...CGAGGCTGTCTTTCATGTCAGGCTGGAGGTC 491

```

```
seq_name: N_Geneseq_36:T31288
```

```

seq_documentation_block:
ID T31288 standard; cDNA; 2919 BP.
AC T31288;
DT 24-FEB-1997 (first entry)
DE Human poly-immunoglobulin receptor, cDNA.
KW Human; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 235..2475
FT /tag= a
FN W09621012-A1.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNWE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.

```


119 rLeuAsnValHisSerGlu 125
:||||:||||| :|||
523 CCTGGAGTGCAGCAAGAT 541

seq_name: N_Geneseq_36:T96033

seq_documentation_block:
ID T96033 standard; cDNA; 2084 BP.

AC T96033;
DE 21-MAY-1998 (first entry)
DE Rat kidney injury related molecule (KIM) cDNA clone 1-7.
KW Kidney injury related molecule; KIM; rat; renal disease; injury;
KW nephritis; tissue regeneration; therapy; ss.
OS Rattus sp.
FH Key Location/Qualifiers
FT CDS 145..1068
FT /tag= a
PN W09744460-A1.
PD 27-NOV-1997.
PE 23-MAY-1997; U09303.
PR 23-AUG-1996; US-023442.
PR 24-MAY-1996; US-018226.
PA (BIOJ) BIOGEN INC.
PI Bonventre JV, Cate RL, Hession CA, Ichimura T, Sanicola-Nadel M,
PI Wei H;
DR WPI; 98-018514/02.
DR P-PSDB; W38334.
PT DNA encoding kidney injury related molecule - which is upregulated
PT in injured or regenerating tissue, useful to promote growth of new
PT tissue and survival of damaged tissue
PS Claim 1; Page 34-37; 68pp; English.
CC cDNA clone 1-7, deposited as ATCC 98060, codes for a rat protein
CC (see W38334), designated kidney injury related molecule (KIM), that
CC is upregulated in injured or regenerating tissue. Representational
CC difference analysis was used to examine cDNA libraries prepared from
CC ischaemic and normal rat adult kidneys. 3 Cycles of hybridisation
CC and selective amplification gave 3 fragments present only in the
CC injured kidney library. Screening of the injured kidney library
CC with these fragments gave clone 1-7, as well as clone 3-2 (see
CC T96033), which is a splice variant of 1-7 encoding an identical KIM
CC protein (nucleotides 136-605 of 3-2 represent an insertion), and
CC clone 4-7 (see T96034) encoding a 307-amino acid KIM protein (see
CC W38335). A labelled probe from clone 1-7 was used to identify a
CC human KIM sequence (see T96035). KIM nucleic acids can be used in
CC the recombinant production of KIM polypeptides in prokaryotic or
CC eukaryotic host cells. KIM, or an agonist, can be used to treat renal
CC disease and to promote the growth of new tissue or the survival of
CC damaged tissue, generally in conditions where the binding of
CC specific ligand to KIM stimulates cell growth, maintains cellular
CC differentiation or reduces apoptosis, e.g. in cases of renal
CC failure, nephritis, kidney transplants, toxic or hypoxic injury.
CC Damage/regeneration of renal cells can be determined by measuring
CC KIM polypeptide or nucleic acid, e.g. with an antisense probe,
CC particularly to diagnose or monitor the progress of disease or
CC therapy.

SQ Sequence 2084 BP; 604 A; 451 C; 470 G; 559 T;

alignment_scores:
Quality: 171.00 Length: 403
Ratio: 0.924 Gaps: 22
Percent Similarity: 45.906 Percent Identity: 24.318

alignment_block:

US-09-135-238B-2 x T96033 ..

Align seg 1/1 to: T96033 from: 1 to: 2084

13 ValSerGlyAlaLeuArgIleLeuPro.....GluVa 23 *
:||||||| ||| :|||||||
166 ATTTCAGGCTCTGCTGCTTCTTCCAGGCTCTGTAGATTCTTATGAAGT 215
23 llysValGluGlyCluLeuGlySerValThrIleLysCysProLeup 40

1 :||||:||||| :||| :||||||| |||
216 A..GTGAAGGGGTGGTGGGTCAACCTGTGCACAAATTCATGTACTACT 262
40 rogluMethHisValArgIleTyLeuCys.....ArgGluMetAlaGly 54
||| ||| :||| :|||
263 CAACACGTGGAGGAATCAACACGACATGTTGGGGCCGGGGCAATGCCCA 312
55 SerGlyThrCysGlyThrValValSerThrThrAsnPhelLysAlaG1 71
:||||:||||| :||| :||||| :||| :||| :|||
313 TATTTCTAGTTGTCAAAATATACTTATTTGGACCAATGGATACCAAGTAC 362
71 uTyLys.....GlyArgValThrLeuLysGlnTyProArgLysAsnL 86
:||||:||||| :||| :||||| :||| :||| :|||
363 CTATCGGAGCAGCGTCGATCAACATAAAGGGCGGTATTTTCAGAAGGAG 412
86 euPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTy 102
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
413 ACGTATCTCTTGACAATAGAGAACTCTGTTGATAGTAGTAGTGTCTGTAT 462
103 AlaCysGlyAla.....GlyMetAsnThrAspArgGlyLysThrG1 116
:||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
463 TGTTCGAGTGAGATTCTCTGGATGTTCAACGAT.....CA 500
116 nLysValThrLeuAsnValHisSerGluTyGluProSerTrpGluGluG 133
:||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
501 GAAATGACCTTTTCATTG.....GAAGTTAAACCAAGAAATCCACAA 544
133 lnProMetProGluThrProLysTrpPheHisLeuProTyLeuPheGln 149
:||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
545 GTCTCT...CCAACAAGACCC..... 561
150 MetProAlaTyAlaSerSerLysPheValThrArgValThrPr 166
:||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
562ACAACTACAGACCCACACACAGGCCCCCAACTAT 599
166 oAlaGlnArgGlyLysValProPro.....ValHisHisSers 179
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
600 TTCAACAAGATCCACACATGTACCAACATCAACAGAGTCTCCACCTCTA 649
179 erProThr.....ThrGlnIleThrHisArgProArgValSerArg 192
:||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
650 CTCACACACGACAGCAACACAG...ACTCACAACCCAGAAATCACT... 693
193 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThrTh 209
:||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
694ACATTTTATGCCCATGAGAC 713
209 rAlaSerLysIleSerAlaLeuGluGlyLeuLysProGlnThrProS 226
:||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
714 AACTGTGAGGTGACA.....GAAACTCCAT 739
226 erTyT..... 227
740 CATATATCTCTGCAGACTGGAATGGCACTGTGCATCTCTCAGAGAGGCC 789
228AsnHisHisThrArgLeuHisArgGlnArgAlaLeuAspTyG1 242
:||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
790 TGGNAATATCACACTGTAAAGATCCCTTTGAGGAG..... 825
242 ySerGlnSerGlyArgGluGlyGlnGlyPheHisIleLeuIleProThrI 259
:||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
826 ...CCGCAGAGAAACCCGACTAAGGGCTTCTATGTTGGCATGTCCGTTG 871
259 leLeuGlyLeuPheLeuLeuAlaLeuLeuGlyLeuValValLysArgAla 275
:||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
872 CAGCCCTGTCTGTCTGCTGCTTCCGAGCACCGGTGGTGTCCACAGGTAC 921
276 ValGluArgGlys.....AlaLeuSerArgArgAlaArgArgLe 289
:||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
922 ATCATATTAAGAAAGAGATGGGCTCTCTGAGCTTTGTGTGCTTCCATGT 971
289 uAlaValArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerP 306
:||||:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```
972 CTCT...AAGAGTAGAGCTTTGCAGACGCA.....GCGATTG 1006
306 roArgProArgSerGlnAsnAsnIleTyrSerAlaCysProArgArgAla 322
   ::::::::::::::::::::::::::::|
1007 TGCATCCCGAGAGCTGAAGACAACACTAC...ATTATTGAAGATAGATCT 1053
   ::::::::::::::::::::|
323 ArgGlyAlaAsp..... 326
   |||
1054 CGAGGTGCAGAGTCCAGAGGCTTCGTGGGGCTTCGTGCTGGG 1103
327 .....AlaAlaGlyThrGlyGluAlaProValProGlyP 338
   ::::|
1104 ATTACAGAGATCGTGACTGATTTCACAGAGTAAATAACCATTCAGCTC 1153
338 roGly.....AlaProLeuProProAlaProLeuGlnValSer 350
   |||||
1154 CTGGGAGATTTGTGTTTGGTTCCTCCAGCTGCAGTGGAGAGGTAACC 1203
351 GluSerPro 353
   |||
1204 CTCTACCCT 1212
```

seq_name: N_Geneseq_36:T96032

seq_documentation_block:

AC T96032 standard; cDNA; 2566 BP.

DT 21-MAY-1998 (first entry)

DE Rat kidney injury related molecule (KIM) cDNA clone 3-2.

KW Kidney injury related molecule; KIM; rat; renal disease; injury;

KW nephritis; tissue regeneration; therapy; ss.

OS Rattus sp.

FX Key Location/Qualifiers

FT CDS 615..1538

ET WPI; 98-018514/02.

FN WO9744460-A1.

PD 27-NOV-1997.

PF 23-MAY-1997; U09303.

PR 23-AUG-1996; US-023442.

PR 24-MAY-1996; US-018228.

FA (BIOJ) BIOGEN INC.

PI Bonventre JV, Cate RL, Hession CA, Ichimura T, Sanicola-Nadel M,

PI Wei H;

DR P-PSDB; W38334.

PT DNA encoding kidney injury related molecule - which is upregulated

PT in injured or regenerating tissue, useful to promote growth of new

PT tissue and survival of damaged tissue

PS Claim 1; Page 32-34; 68pp; English.

CC (see W38334), designated as ATCC 98061, codes for a rat protein

CC is up-regulated in injured or regenerating tissue. Representational

CC difference analysis was used to examine cDNA libraries prepared from

CC ischaemic and normal rat adult kidneys. 3 Cycles of hybridisation

CC and selective amplification gave 3 fragments present only in the

CC injured kidney library. Screening of the injured kidney library

CC with these fragments gave clone 3-2, as well as clone 1-7 (see

CC T96033), which is a splice variant of 3-2 encoding an identical KIM

CC polypeptide (nucleotides 136-605 of 3-2 represent an insertion),

CC and clone 4-7 (see T96034) encoding a 307-amino acid KIM protein

CC (see W38335). Clone 1-7 was used to identify a human KIM sequence

CC (see T96035). KIM nucleic acids can be used in the recombinant

CC production of KIM polypeptides in prokaryotic or eukaryotic host

CC cells. KIM, or an agonist, can be used to treat renal disease and to

CC promote the growth of new tissue or the survival of damaged tissue,

CC generally in conditions where the binding of specific ligand to KIM

CC stimulates cell growth, maintains cellular differentiation or

CC reduces apoptosis, e.g. in cases of renal failure, nephritis,

CC kidney transplants, toxic or hypoxic injury. Damage/regeneration

CC of renal cells can be determined by measuring KIM polypeptide or

CC nucleic acid, e.g. with an antisense probe, particularly to

CC diagnose or monitor the progress of disease or therapy.

CC Sequence 2566 BP; 726 A; 546 C; 591 G; 703 T;

SQ

```
alignment_scores:
  Quality: 171.00      Length: 403
  Ratio: 0.924        Gaps: 22
  Percent Similarity: 45.906      Percent Identity: 24.318

alignment_block:
US-09-135-238B-2 x T96032 ..
Align seg 1/1 to: T96032 from: 1 to: 2566
13 ValSerGlyAlaLeuArgIleLeuPro.....GluVa 23
   :::||||| ||| ::::|
636 ATTTCAGGCTCTCTGCTGCTTCTCCAGGCTGTGTAGATCTTATGAAGT 685
23 lLysValGluGlyGluLeuGlySerValThrIleLysCysProLeuP 40
   | |||::| |::| | |||||
686 A...GTGAAGGGGTGGTGGGTCACTGTCACAATTCCATGTACTTACT 732
40 roGluMetHisValArgIleTyrLeuCys.....ArgGluMetAlaGly 54
   ||| ||| ::::|
733 CAACAGCTGGAGGAATCAACAACGACATGTTGGGGCCGGGCAATGCCCA 782
55 SerGlyThrCysGlyThrValValSerThrThrAsnPhelIleLysAlaGl 71
   ::::| ||| ::::| | |||||
783 TATCTAGTTGTTCAAAATATACTTATTGGACCAATGGATACCAAGTCAC 832
71 uTyrLys.....GlyArgValThrLeuLysGlnTyrProArgLysAsnL 86
   ||::| | ||||| ::::| |
833 CTATCGAGCAGCGGTCGATACACAATAAAGGGCGGTATTTCAGAAGGAG 882
86 euPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyr 102
   ::::| |::| |::| | |||||
883 ACGTATCTCTGACAATAGAGACTCTGTTGATAGTAGTATGCTGTGTAT 932
103 AlaCysGlyAla.....GlyMetAsnThrAspArgGlyLysThrGl 116
   ::::| |::| | |||||
933 TGTTCGGAGTGGAGATTCTGTTGATGTTCAAGCAT.....CA 970
116 nLysValThrLeuAsnValHisSerGlyTyrGluProSerTrpGluGluG 133
   ||||| |::| |::| | |||||
971 GAAATGACCTTTTCATTG.....GAAGTTAAACCAGAAATCCCAACA 1014
133 lnProMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGln 149
   ||||| ||| |||
1015 GTCTT...CCAAACAAGACCC..... 1031
150 MetProAlaTyrAlaSerSerLysPheValThrArgValThrThrPr 166
1032 .....ACAACTACAGACCCCAACCAAGGCCCAACTAT 1069
166 oAlaGlnArgGlyLysValProPro.....ValHisHisSerS 179
   ::| ||::| | |||
1070 TTCAACAAGATCCACACATGTACCAACATCAACAGAGTCTCCACCTCA 1119
179 erProThr.....ThrGlnIleThrHisArgProArgValSerArg 192
   ||||| ||||| | |||||
1120 CTCCAACACCAAGAACAAACACAG...ACTCAAAACACAGAAATCACT... 1163
193 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThrTh 209
1164 .....ACATTTTATGCCCATGAGAC 1183
209 rAlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProGlnThrProS 226
   |::|::|::|::|
1184 AACTGCTGAGGTGACA.....GAACTCCAT 1209
226 erTyr..... 227
   |||||
1210 CATATCTCTCTGAGACTGGAATGGCACTGTGACATCTCAGAGGAGGCC 1259
228 .....AsnHisThrArgLeuHisArgGlnArgAlaLeuAspTyrGl 242
```


Tue Sep 12 08:57:54 2000

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2000, 10:09:19 ; Search time 88.29 Seconds
(without alignments)
5412.469 Million cell updates/sec

Title: US-09-135-238B-1

Perfect score: 1910
Sequence: 1 aaaggaagacgagcgtgctc.....ttactctgtctccatcctt 1910

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq.36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1830.4	95.8	2040	1 X28178	Human PIGRL-1 codi
2	676	35.4	1047	1 X28179	Human PIGRL-1 codi
3	45.2	2.4	117213	1 V62176	HSV-2 strain SB5 C
4	41	2.1	1911	1 V20383	cDNA for human imm
5	40.4	2.1	1984	1 V19114	Human secreted apo
6	40.2	2.1	987	1 T89157	Oerskovia xanthine
7	40.2	2.1	1516	1 T89156	Oerskovia xanthine
8	39.6	2.1	3850	1 X21656	Human helicase gen
9	39	2.0	117213	1 V62176	HSV-2 strain SB5 C
10	38.6	2.0	1512	1 T20346	Chicken yesc proto-
11	38.6	2.0	114955	1 X53491	Human adenosine A1
12	38.2	2.0	1593	1 T29199	Rat acetylglucosam
13	38.2	2.0	2247	1 Q57973	Human glycosyltran
14	37.4	2.0	429	1 Q69901	rAb Stage 1 VH con
15	37.4	2.0	2094	1 T33198	Alpha-1 A/D adrena
16	37	1.9	659	1 V15999	NBCCS (PTC) protei
17	37	1.9	3089	1 T05572	MEK33 cDNA. New si
18	37	1.9	3089	1 V22678	cDNA encoding a mi
19	37	1.9	3089	1 X5616	MEK33 protein codi
20	37	1.9	114955	1 X53491	Human adenosine A1
21	36.8	1.9	4258	1 Q22439	DNA of hEco-1, en
22	36.8	1.9	4258	1 T66529	Human C-type natri
23	36.8	1.9	6225	1 X55273	Human enzyme-relat
24	36.6	1.9	3841	1 T79967	Presenilin-interac
25	36.6	1.9	15872	1 T68715	Streptomyces venez
26	36.4	1.9	1417	1 V27893	Human hypothalamia
27	36.4	1.9	1929	1 Q43590	VCAM-6D/ICAM-1 DNA
28	36.4	1.9	1932	1 Q46661	VCAM-6D/ICAM-2 DNA
29	36.4	1.9	1941	1 Q43586	VCAM-6D DNA. Monoc
30	36.4	1.9	1941	1 Q46662	VCAM-6D/VCAM4-1 DN
31	36.4	1.9	2091	1 Q23321	Soluble VCAM varia
32	36.4	1.9	2205	1 Q43587	VCAM/ICAM-1 DNA. M
33	36.4	1.9	2208	1 Q43588	VCAM/ICAM-2 DNA. M

34 36.4 1.9 2217 1 Q43585
35 36.4 1.9 2220 1 Q10205
36 36.4 1.9 2220 1 Q21005
37 36.4 1.9 2220 1 Q29043
38 36.4 1.9 2220 1 Q43589
39 36.4 1.9 2220 1 Q43394
40 36.4 1.9 2220 1 V16209
41 36.4 1.9 2811 1 Q06687
42 36.4 1.9 3080 1 Q06688
43 36.4 1.9 3975 1 N81157
44 36.2 1.9 3975 1 Q22999
45 36.2 1.9 4020 1 T91361

VCAM-7D DNA. Monoc
Sequence encoding
VCAM variant with
1E7-2G7 antigen cD
VCAM/ICAM-3 DNA. M
Sequence of bases
cDNA encoding huma
Vascular cell adhe
Vascular cell adhe
Malaria-specific g
SERP gene. Recombi
Orf virus genomic

ALIGNMENTS

RESULT 1
X28178
ID X28178 standard; cDNA; 2040 BP.
AC X28178;
DT 16-JUN-1999 (first entry)
DE Human PIGRL-1 coding sequence.
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;
KW X-linked Severe Combined Immunodeficiency; XSCID; Iga deficiency;
OS Homo sapiens.
PN EP-905238-A2.
PD 31-MAR-1999.
PF 14-AUG-1998; 306487.
PR 30-OCT-1997; US-961564.
PR 25-OCT-1997; US-056935.
PA (SMK) SMITHKLINE BEECHAM CORP.
PI Sweet RW, Truneh A, Wu S;
DR WPI: 99-192666/17.
DT P-PSDB; Y05001.
PT New polypeptides encoding PIGRL-1 useful for treating diseases such
PT as X-linked Severe Combined Immunodeficiency
PS Claim 4; Page 18-19; 26pp; English.
CC This sequence encodes the human PIGRL-1 protein of the invention.
CC Autoimmune diseases involving altered expression or activity of PIGRL-1
CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined
CC Immunodeficiency (XSCID) and Iga deficiency. These diseases can be
CC diagnosed or susceptibility to them predicted by: (1) determining whether
CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or
CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.
CC Patients deficient in PIGRL-1 can be treated by administering either the
CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.
CC Patients with excessive expression or activity of PIGRL-1 can be treated
CC by administering an antagonist of PIGRL-1, an antisense nucleic acid
CC molecule which inhibits the expression of PIGRL-1 or administering
CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can
CC be used to identify its agonists by contacting a cell expressing PIGRL-1
CC with a candidate compound in the presence of a signal system and noting
CC the candidate as an agonist if a signal is produced. The same method can
CC be used to identify antagonists of PIGRL-1 but the presence of an
CC antagonist is indicated by a decrease in production of the signal.
CC Antibodies against PIGRL-1 may be used to isolate or identify clones
CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to
CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.
CC This information may then be correlated with the incidence of autoimmune
CC disease in those patients to identify whether the mutation causes the
CC disease.
SQ Sequence 2040 BP; 478 A; 616 C; 494 G; 452 T;

Query Match 95.8%; Score 1830.4; DB 1; Length 2040;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1893; Conservative 0; Mismatches 11; Indels 6; Gaps 5;
QY 5 agatgaagcagcgtgctccatccctctctagggctcttgagccttgacactcactca 64
77 GAGTAAGCAGCGTGCTCCATCCCTCTCTAGGGGCTCTTGGAGGACCTTGCACACTCA 136

PT as X-linked Severe Combined Immunodeficiency

PS Disclosure; Page 8; 2pp; English.
 CC This sequence encodes the human PIGRL-1 protein of the invention.
 CC Autoimmune diseases involving altered expression or activity of PIGRL-1
 CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined
 CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be
 CC diagnosed or susceptibility to them predicted by: (1) determining whether
 CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or
 CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.
 CC Patients deficient in PIGRL-1 can be treated by administering either the
 CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.
 CC Patients with excessive expression or activity of PIGRL-1 can be treated
 CC by administering an antagonist of PIGRL-1, an antisense nucleic acid
 CC molecule which inhibits the expression of PIGRL-1 or administering
 CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can
 CC be used to identify its agonists by contacting a cell expressing PIGRL-1
 CC with a candidate compound in the presence of a signal system and noting
 CC the candidate as an agonist if a signal is produced. The same method can
 CC be used to identify antagonists of PIGRL-1 but the presence of an
 CC antagonist is indicated by a decrease in production of the signal.
 CC Antibodies against PIGRL-1 may be used to isolate or identify clones
 CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to
 CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.
 CC This information may then be correlated with the incidence of autoimmune
 CC disease in those patients to identify whether the mutation causes the
 CC disease.
 SQ Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T;

Query Match 35.4%; Score 676; DB 1; Length 1047;
 Best Local Similarity 94.6%; Pred. No. 8e-176;
 Matches 785; Conservative 0; Mismatches 34; Indels 11; Gaps 8;

QY 5 gagtaagcagcgtgtctccatcccccctctctagggccttggatggacccttgcaactcta 64
 DB 81 GAGTAAGCAGCGTGTCTCCATCCCTCTCTAGGGCTCTTGGATGGACCTTGCACCTTA 140
 QY 65 gaaaggacaatggactctggcttggccactttacttctctccagttatcaggggccctg 124
 DB 141 GAAGGACAATGGACTTCTGGCTTGGCCACTTTACTTCTGCCAGTATCAGGGGCCCTG 200
 QY 125 aggatctcccaagaagtaaaagtagaggggagactggcgagtcagttaccatcaaatgc 184
 DB 201 AGGATCTCCCAAGAGTAAGAGGTAGAGGGGAGCTGGCGGATCAGTTACCATCAAGTGC 260
 QY 185 ccacttctgaaatgcagtgtgagatatactgtgccggagatggtggatctggaaca 244
 DB 261 CCACTTCTGAAATGCAATGTGAGGATATATCTGTGCCGGAGATGGCTGGATCTGGAACA 320
 QY 245 tctgtaccgtgtatccaccacacacttcacatcaagcagaatacaaggccgagttact 304
 DB 321 TGTGTTACCTGGTATCCACCACTTCATCAAGCAGAGATACAGGGCCGAGTTACT 380
 QY 305 ctgaagcaataccaccacgaagaatctgttcttagtggaggttaacacagctgcagaaagt 364
 DB 381 CTGAAGCAATAACCCACGCAAGATCTGTCTAGTGGAGGTAAACACAGCTGCAGAAAAGT 440
 QY 365 gacagcggagtctatgctcgcagcggcgatgaacacagacccgggaaagaccagaaa 424
 DB 441 GACAGCGAGTCTATGCTCTCGGA-CGGGCTGAAACACACAGACCGGGGAAAGACCCAGAAA 499
 QY 425 gtacacctgaatgtccacagtgaatacagaccatcatgtggaagagacagcgaatgcctgag 484
 DB 500 GTACACCTGAATGTCCACAGTGAATACAGCCATCATGGGAAGAGCAGCAATGCCTGAG 559
 QY 485 actcaaaatggtttcctatccttctgttccagatccctgcataatgcagttctcc 544
 DB 560 ACTCAAAATGTTTATCTGTCCCTATTGTTCCAGATGCCCTGCATATGCCGGTTCTTCC 619
 QY 545 aaattcgttaac--cagagttaccacacacagc-tcaaggggcaaggttcctccagttcac 601
 DB 620 ACATTCTGTAACCGCAGGATTACACACACAGCTTCAAGAGGGCAAGGTCCCTCCAGTTTAC 679

QY 602 caactctccccccaccaccacaaa-tcaccaccgccc-tcagtgctccagagcattcttcag 659
 DB 680 CACTCTCTCCCCACCACCACCAATTCACCCAGCCCTTCGAGTGTNCAGAGCATCTTCAG 739
 QY 660 tagcaggtgacagcccg-aacctctctgcatccactacagcctcaaaaatctcagct 718
 DB 740 TAGCAGGTGACAGCCCGAATCTTCTGTCATCCACTAGAGCTCAAAAATCTCAGCT 799
 QY 719 ctggaggggctctcaagccccc--agcgcgccaggtctacaaccacacacagcgtgcac 775
 DB 800 CTGGAAGGCTCTTCAAGCCCGCAGAGCGCCAGCTACAA-CANCAACACAGGCTGCAC 858
 QY 776 aggcagagagcctgagctactatgctccacagctctggaggaaggccaaag 825
 DB 859 AGGCAGAGAGCACTGGATATTATGGGNTACAGCTCTGGGAGGGGAANG 908

RESULT 3
 V62176
 ID V62176 standard; DNA; 117213 BP.
 AC V62176;
 DT 13-JAN-1999 (first entry)
 DE HSV-2 strain SB5; immunological response induction; therapy;
 KW HSV-2 strain SB5; immunological response induction; therapy;
 KW antiviral identification; viral protein inhibitor; ss.
 OS Herpes simplex virus type 2.
 FH Location/Qualifiers
 Key 755..1297
 CDS
 FT /tag= a
 FT /product= "ORF#1 protein"
 FT /note= "encoded protein shown in W72170"
 FT 1170..2174
 FT /tag= b
 FT /product= "ORF#2 protein"
 FT /note= "encoded protein shown in W72171"
 FT 2229..2930
 FT /tag= c
 FT /product= "ORF#3 protein"
 FT /note= "encoded protein shown in W72172"
 FT complement (3130..3735)
 FT /tag= d
 FT /product= "ORF#4 protein"
 FT /note= "encoded protein shown in W72173"
 FT complement (3802..6447)
 FT /tag= e
 FT /product= "ORF#5 protein"
 FT /note= "encoded protein shown in W72174"
 FT 6017..8482
 FT /tag= f
 FT /product= "ORF#6 protein"
 FT /note= "encoded protein shown in W72250"
 FT 6026..8482
 FT /tag= g
 FT /product= "ORF#6f protein"
 FT /note= "encoded protein shown in W72249"
 FT 6065..8482
 FT /tag= h
 FT /product= "ORF#6e protein"
 FT /note= "encoded protein shown in W72248"
 FT 6167..8482
 FT /tag= i
 FT /product= "ORF#6d protein"
 FT /note= "encoded protein shown in W72247"
 FT 6296..8482
 FT /tag= j
 FT /product= "ORF#6c protein"
 FT /note= "encoded protein shown in W72246"
 FT 6326..8482
 FT /tag= k
 FT /product= "ORF#6b protein"
 FT /note= "encoded protein shown in W72245"
 FT 6446..8482
 FT /tag= l
 FT

```
FT /product= "ORF#6a protein"
FT /transl_except= (pos: 7400..7402, aa: Ala-Ala)
FT /transl_except= (pos: 7481..7486, aa: Ile)
FT /note= "encoded protein shown in W72244"
FT 8457..9347
CDS
FT /*tag= m
FT /product= "ORF#7 protein"
FT /note= "encoded protein shown in W72175"
FT complement (9604..11855)
FT /*tag= n
CDS
FT /transl_except= (pos: 11635..11636, aa: Ala)
FT /product= "ORF#8 protein"
FT /note= "encoded protein shown in W72176"
FT complement (11905..14508)
FT /*tag= o
CDS
FT /product= "ORF#9b protein"
FT /note= "encoded protein shown in W72222"
FT complement (11905..14520)
FT /*tag= p
CDS
FT /product= "ORF#9a protein"
FT /note= "encoded protein shown in W72223"
FT 14399..15802
FT /*tag= q
FT /product= "ORF#10 protein"
FT /note= "encoded protein shown in W72177"
FT complement (15996..16286)
FT /*tag= r
CDS
FT /product= "ORF#11 protein"
FT /note= "encoded protein shown in W72178"
FT complement (16202..18064)
FT /*tag= s
FT /product= "ORF#12 protein"
FT /note= "encoded protein shown in W72179"
FT complement (18105..19861)
FT /*tag= t
CDS
FT /product= "ORF#13 protein"
FT /note= "encoded protein shown in W72180"
FT complement (19415..20074)
FT /*tag= u
CDS
FT /product= "ORF#14 protein"
FT /note= "encoded protein shown in W72181"
FT 20155..21453
FT /*tag= v
FT /product= "ORF#15 protein"
FT /note= "encoded protein shown in W72182"
FT complement (21326..22291)
FT /*tag= w
CDS
FT /product= "ORF#16 protein"
FT /note= "encoded protein shown in W72183"
FT complement (22546..24654)
FT /*tag= x
FT /product= "ORF#17 protein"
FT /note= "encoded protein shown in W72184"
FT 24684..25955
FT /*tag= y
FT /product= "ORF#18 protein"
FT /note= "encoded protein shown in W72185"
FT complement (26295..27251)
FT /*tag= z
CDS
FT /product= "ORF#19 protein"
FT /note= "encoded protein shown in W72186"
FT complement (27630..31754)
FT /*tag= aa
FT /product= "ORF#20b protein"
FT /note= "encoded protein shown in W72225"
FT complement (27630..31784)
FT /*tag= ab
CDS
FT /product= "ORF#20a protein"
FT /note= "encoded protein shown in W72224"
FT complement (32067..32735)
FT /*tag= ac
FT /product= "ORF#21 protein"
FT /note= "encoded protein shown in W72187"
FT 33140..34984
FT /*tag= ad
FT /product= "ORF#22a protein"
FT /note= "encoded protein shown in W72226"
FT 33386..34984
CDS
FT /*tag= ae
FT /product= "ORF#22b protein"
FT /note= "encoded protein shown in W72227"
FT complement (35205..37721)
FT /*tag= af
CDS
FT /product= "ORF#23 protein"
FT /note= "encoded protein shown in W72188"
FT complement (38058..39188)
FT /*tag= ag
CDS
FT /product= "ORF#24 protein"
FT /note= "encoded protein shown in W72189"
FT 39090..39935
FT /*tag= ah
FT /product= "ORF#25 protein"
FT /note= "encoded protein shown in W72190"
FT 40216..41973
FT /*tag= ai
CDS
FT /product= "ORF#26 protein"
FT /note= "encoded protein shown in W72191"
FT 42206..44178
FT /*tag= aj
FT /transl_except= (pos: 44063..44064, aa: Lys)
FT /product= "ORF#27 protein"
FT /note= "encoded protein shown in W72192"
FT complement (44853..47297)
FT /*tag= ak
CDS
FT /product= "ORF#28 protein"
FT /note= "encoded protein shown in W72193"
FT 47122..47338
FT /*tag= al
FT /product= "ORF#29 protein"
FT /note= "encoded protein shown in W72194"
FT complement (47305..49662)
FT /*tag= am
CDS
FT /product= "ORF#30 protein"
FT /note= "encoded protein shown in W72195"
FT complement (50035..51666)
FT /*tag= an
CDS
FT /product= "ORF#31 protein"
FT /note= "encoded protein shown in W72196"
FT complement (51701..53575)
FT /*tag= ao
FT /product= "ORF#32 protein"
FT /note= "encoded protein shown in W72197"
FT 54393..58115
FT /*tag= ap
CDS
FT /product= "ORF#33 protein"
FT /note= "encoded protein shown in W72198"
FT complement (58060..58977)
FT /*tag= aq
FT /product= "ORF#34 protein"
FT /note= "encoded protein shown in W72199"
FT complement (58970..60760)
FT /*tag= ar
CDS
FT /product= "ORF#35 protein"
FT /note= "encoded protein shown in W72200"
FT 60759..61151
FT /*tag= as
FT /product= "ORF#36 protein"
FT /note= "encoded protein shown in W72201"
FT 61241..62071
FT /*tag= at
CDS
FT /product= "ORF#37 protein"
FT /note= "encoded protein shown in W72202"
FT 62183..62521
FT /*tag= au
FT /product= "ORF#38 protein"
FT /note= "encoded protein shown in W72203"
FT 62183..62521
```

ID V19114 standard; DNA; 1984 BP.
 AC V19114;
 DT 28-AUG-1998 (first entry)
 DE Human secreted apoptosis-related protein hSARP3 DNA.
 KW Secreted apoptosis-related protein; SARP; hSARP; human;
 KW prostate cancer; breast cancer; diagnosis; gene therapy; ds.
 OS Homo sapiens.

FR 24-SEP-1997; 017-0264363.
PT 11-OCT-1996; US-028363.
PR 24-SEP-1996; US-026603.
PA (LXRB-) LAR BIOTECHNOLOGY INC.
PI MelKonyan H, Umansky S;
PI WPI: 98-230704/20.
DR P-PSDB; W37816.
PT P-PTN; W37816.
PT apoptosis, particularly for treatment of prostatic or breast cancer,
PT also for diagnosis and monitoring of disease
PS Claim 2; Page 54-56; 10pp; English.
CC This nucleic acid sequence encodes human secreted apoptosis-related
CC protein hSARP3 (see W37816) that modulates apoptosis through

Query Match 2.1%; Score 40.4; DB 1; Length 1984;
Best Local Similarity 46.0%; Pred. No. 0.2;
Matches 137; Conservative 0; Mismatches 161; Indels 0; Gaps

	RESULT	5
QY	ccgctgcagggtctcgaatccccctggctccatccccctcttcgaagaccagtcttg	1168
1108	ccgctgcagggtctcgaatccccctggctccatccccctcttcgaagaccagtcttg	1163


```

FT      /tag= t
FT      /product= "ORF#13 protein"
FT      /note= "encoded protein shown in W72180"
FT      complement (19415..20074)
FT      /tag= u
FT      /product= "ORF#14 protein"
FT      /note= "encoded protein shown in W72181"
FT      20155..21453
FT      /tag= v
FT      /product= "ORF#15 protein"
FT      /note= "encoded protein shown in W72182"
FT      complement (21326..22291)
FT      /tag= w
FT      /product= "ORF#16 protein"
FT      /note= "encoded protein shown in W72183"
FT      complement (22546..24654)
FT      /tag= x
FT      /product= "ORF#17 protein"
FT      /note= "encoded protein shown in W72184"
FT      24684..25955
FT      /tag= y
FT      /product= "ORF#18 protein"
FT      /note= "encoded protein shown in W72185"
FT      complement (26295..27251)
FT      /tag= z
FT      /product= "ORF#19 protein"
FT      /note= "encoded protein shown in W72186"
FT      complement (27630..31754)
FT      /tag= aa
FT      /product= "ORF#20b protein"
FT      /note= "encoded protein shown in W72225"
FT      complement (27630..31784)
FT      /tag= ab
FT      /product= "ORF#20a protein"
FT      /note= "encoded protein shown in W72224"
FT      complement (32067..32735)
FT      /tag= ac
FT      /product= "ORF#21 protein"
FT      /note= "encoded protein shown in W72187"
FT      33140..34984
FT      /tag= ad
FT      /product= "ORF#22a protein"
FT      /note= "encoded protein shown in W72226"
FT      33386..34984
FT      /tag= ae
FT      /product= "ORF#22b protein"
FT      /note= "encoded protein shown in W72227"
FT      complement (35205..37721)
FT      /tag= af
FT      /product= "ORF#23 protein"
FT      /note= "encoded protein shown in W72188"
FT      complement (38058..39188)
FT      /tag= ag
FT      /product= "ORF#24 protein"
FT      /note= "encoded protein shown in W72189"
FT      39090..39935
FT      /tag= ah
FT      /product= "ORF#25 protein"
FT      /note= "encoded protein shown in W72190"
FT      40216..41973
FT      /tag= ai
FT      /product= "ORF#26 protein"
FT      /note= "encoded protein shown in W72191"
FT      42206..44178
FT      /tag= aj
FT      /transl_except= (pos: 44063..44064, aa: Lys)
FT      /product= "ORF#27 protein"
FT      /note= "encoded protein shown in W72192"
FT      complement (44853..47297)
FT      /tag= ak
FT      /product= "ORF#28 protein"
FT      /note= "encoded protein shown in W72193"
FT      47122..47338

```

```

FT      /tag= al
FT      /product= "ORF#29 protein"
FT      /note= "encoded protein shown in W72194"
FT      complement (47305..49662)
FT      /tag= am
FT      /product= "ORF#30 protein"
FT      /note= "encoded protein shown in W72195"
FT      complement (50035..51666)
FT      /tag= an
FT      /product= "ORF#31 protein"
FT      /note= "encoded protein shown in W72196"
FT      complement (51701..53575)
FT      /tag= ao
FT      /product= "ORF#32 protein"
FT      /note= "encoded protein shown in W72197"
FT      54393..58115
FT      /tag= ap
FT      /product= "ORF#33 protein"
FT      /note= "encoded protein shown in W72198"
FT      complement (58060..58977)
FT      /tag= aq
FT      /product= "ORF#34 protein"
FT      /note= "encoded protein shown in W72199"
FT      complement (58970..60760)
FT      /tag= ar
FT      /product= "ORF#35 protein"
FT      /note= "encoded protein shown in W72200"
FT      60759..61151
FT      /tag= as
FT      /product= "ORF#36 protein"
FT      /note= "encoded protein shown in W72201"
FT      61241..62071
FT      /tag= at
FT      /product= "ORF#37 protein"
FT      /note= "encoded protein shown in W72202"
FT      62183..62521
FT      /tag= au
FT      /product= "ORF#38 protein"
FT      /note= "encoded protein shown in W72203"

```

Query Match 2.08; Score 39; DB 1; Length 117213;

Best Local Similarity 50.8%; Pred. No. 3.5; Matches 93; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

```

QY 924 ggcgggcccgcgactgagatgcgcgcctggagagctccagagagcccgccg 983
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114214 gccgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 114155

QY 984 ggtcgccgacgcgcctcccaaacatctacagcctcccgccgcgcctctgg 1043
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114154 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 114095

QY 1044 agcggacgctgcaggcacagggaggcccccttcccgccccggagcgcttgc 1103
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114094 tcctgacacgcgtctgaggggggggggggggggggggggggggggggg 114035

QY 1104 cgc 1106
      ||
Db 114034 gcc 114032

```

RESULT 10

T30346/c

ID T30346 standard; cDNA; 1512 BP.

AC T30346;

DT 30-AUG-1996 (first entry)

DE Chicken Yes proto-oncogene associated protein YAP65 cDNA.

KW Yes proto-oncogene associated protein; YAP; YAP65;

KW signal transduction; cancer; muscular dystrophy; WW domain;

KW diagnosis; gene therapy; ds.

OS Gallus sp.

FT Key Location/Qualifiers

FT cds 66..1412

[illegible]

```

OS
FH Key Location/Qualifiers
FT signal_peptide 1..57
FT /*tag= a
FT 58..429
FT mat_peptide /*tag= b
FT
PN W09416094.A.
PD 21-JUL-1994.
PD 07-JAN-1994; U00266.
PR 12-JAN-1993; US-004798.
PR (BIOJ ) BIOGEN INC.
PI Carr EJ, Lobb RR, Tempest PR;
PI WPI: 94-249240/30.
DR P-PSDB: R5940.
PT Recombinant and humanised anti-VLA4 antibodies - contg.
PT non-human CDRs in their light and heavy chains, used to treat
PT (non-)specific inflammation, e.g. asthma
PS Example 3: Page 90: 140pp: English.
PS

```

This sequence encodes the stage I heavy chain variable region (VH) of the recombinant antibody (rAb) of the invention. The VH sequence has been CDRI grafted and comprises a NEMM framework and CDRs derived from the murine monoclonal antibody, Hpl/2. In addition to CDR grafting, the stage I VH construct contains selected framework changes. Just prior to CDRI, a block of residues has been changed to murine residues, Phe28, Asn38, Ile29 and Lys30. Although these four residues are not non-human residues included in CDRI, structurally they are included in part of the CDRI and are empirically included as part of CDRI. Arg was also changed to Asp at position 94. This sequence was used in the production of a humanised recombinant anti-VIA4 antibody (rAb). The humanised rAb produced comprises non-human CDRs at VH positions 31-35 (CDR1), 50-55 (CDR2) and 95-102 (CDR3). It may further comprise non-human residues at framework positions 27-30, 75, 77-79 or 66-67, or 69-71 and 69-71, or 84-85, or 38 and 40 or 24. The VL of the rAb has non-human CDRs at positions 24-34 (CDRI), 50-56 (CDR2) and 89-97 (CDR3). It may also have non-human residues at framework positions 60 and 67. The non-human residues are derived from Hpl/2, which is a member of Kabat V with no unusual residues. The rAb can be used to treat inflammation in mammals, esp. asthma or inflammatory bowel disease.

Example 3: Page 90; 140pp; English.
 This sequence encodes the stage I heavy chain variable region (VH) of the recombinant antibody (rAb) of the invention. The VH sequence has been CDRI grafted and comprises a NEMM framework and CDRs derived from the murine monoclonal antibody, Hpl/2. In addition to CDR grafting, the stage I VH construct contains selected framework changes. Just prior to CDRI, a block of residues has been changed to murine residues, Phe28, Asn38, Ile29 and Lys30. Although these four residues are not non-human residues included in CDRI, structurally they are included in part of the CDRI and are empirically included as part of CDRI. Arg was also changed to Asp at position 94. This sequence was used in the production of a humanised recombinant anti-VIA4 antibody (rAb). The humanised rAb produced comprises non-human CDRs at VH positions 31-35 (CDR1), 50-55 (CDR2) and 95-102 (CDR3). It may further comprise non-human residues at framework positions 27-30, 75, 77-79 or 66-67, or 69-71 and 69-71, or 84-85, or 38 and 40 or 24. The VL of the rAb has non-human CDRs at positions 24-34 (CDRI), 50-56 (CDR2) and 89-97 (CDR3). It may also have non-human residues at framework positions 60 and 67. The non-human residues are derived from Hpl/2, which is a member of Kabat V with no unusual residues. The rAb can be used to treat inflammation in mammals, esp. asthma or inflammatory bowel disease.

Sequence 429 BP: 99 A; 115 C; 136 G; 89 T
 Sequence

```
Query Match      2.0%; Score 37.4; DB 1; Length 429;
Best Local Similarity 53.8%; Pred. No. 0.64; 56; Indels 0; Gaps 0;
Matches 77; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 285 aatacaaggccgagttactctgaagcaataaccacacagaatctgttcttagtgaggg 344
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 245 AGTTCAGGCTCAGAGTGACAAATGCTGTGTAGACACGACGACGACGACGACGACGAC 304
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 345 taacacagctgcagaaagtagacagcgagctctatgctgcggagcgagcgatgaacacag 404
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 305 TCAGAGCCTGACGACCGCCGACCGCGGCTATATTGTGACAGCAATGFGGTAT 364
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 405 accggggaagaccagaaagtc 427
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 365 CAACGGGATATGCTCTGGACTTC 387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
T33198/c
ID T33198 standard; cDNA to mRNA; 2094 BP.
AC T33198;
DT 06-NOV-1996 (first entry)
DE Alpha-1 A/D adrenaline receptor coding sequence.
KW Alpha-1 A/D adrenaline receptor; agonist; antagonist; detection;
KW screening; recombinant production; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 22..1740
FT /*tag= a
FT /product= alpha-1 A/D adrenaline receptor
PN J07289265-A.
PD 07-NOV-1995.
PF 28-APR-1994; 091041.
PR 28-APR-1994; JP-091041.
PA (NNSH ) NIPPON SHINYAKU CO LTD.
DR WPI; 96-015270/02.
DR P-PDB; R98563.
PT DNA encoding alpha-1A/D adrenaline receptor - useful for
PT identification of (ant)agonist(s)
PS Disclosure; ; lpp; Japanese.
CC The present sequence encodes a human alpha-1 A/D adrenaline receptor.
CC The DNA can be used to express the receptor in large quantities by
CC recombinant DNA technology. (Ant)agonists selective to the receptor
CC can also be detected.
SQ Sequence 2094 BP; 305 A; 726 C; 707 G; 356 T;

Query Match      2.0%; Score 37.4; DB 1; Length 2094;
Best Local Similarity 48.4%; Pred. No. 1.4;
Matches 104; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 915 cccctccagggcgccgcgactggcgtgagtgatgcgcccctggagagctcccaga 974
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 238 CCGGCTCCCCCGGAGGCTCCGGTGTCTCGCCGCTGCTGCGCCACGACGCGCGCGC 179
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 975 ggcctcgaggcgccgacgcgctcccaaaacacatctacagcgcttcccgccg 1034
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178 CGCGCGCCCGCCCGCCGGAACCGCGCCACCGCGCGGCGCTCGAGGGGCGCGCGCGC 119
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1035 gcgctctgagcgagcgagcgagcgagggagggcccgcttcccgcccgagcgcc 1094
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 118 CCGCGCTGCCCCCGCGCGCGCGCTGGAGCCCGCTGCGCTGCTGCGGGGGGTC 59
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1095 gttgcccccccgcccgctgaggtgtctgaatctc 1129
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 58 CCTCGAAACTGACGCTCAGGAGATCGCGGAAAGTC 24
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: September 12, 2000, 11:58:00
Job time: 6521 sec

17 uAlaValArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerP 34
|||||
411 GCGCTGAGGATGCGCGCCCTGGAGAGCTCCAGAGCGCCGCGGTCGC 362
34 roArgProArgSerGlnAsnAsnLleYrSerAlaCysProArgArgAla 50
|||||
361 CGCGACCGCGCTCCCAAAACAATCTACAGCGCTTGCCGCGCGCGCT 312
51 ArgGlyAlaAspAlaAlaGlyThrGlyGluAlaProValProGlyProG1 67
|||||
311 CGTGAGCGGAGCGCTGCAGGCACAGGAGGAGCGCCGCTTCGCGCCCGG 262
67 yAlaProLeuProProAlaProAlaProLeuGlnValSerGluSerProfrpLeuH 84
|||||
261 AGCGCGCTTGCGCCCGCGCGCGCTGCAGGTGTTGAATCTCCCTGGCTCC 212
84 isAlaProSerLeuLysThrSerCysGluTyYrValSerLeuTyYrHisGln 100
|||||
211 ATGCCCATCTCTGAGACCACTGCTGAATACGTGAGCGCTTACCACCAG 162
101 ProAlaAlaMetMetGluAspSerAspSerAspAspTyYrIleAsnValPr 117
|||||
161 CTGCGCGCATGATGGAGGACAGTGAATCAGATGACTACATCAATGTTC 112
117 oAla 118
|||||
111 TGCC 108
seq_name: gb_est7:AA456707
seq_documentation_block: 404 bp mRNA EST 06-JUN-1997
LOCUS AA456707 2313904.r1 Soares_NHMPu_S1 Homo sapiens CDNA clone IMAGE:813174
DEFINITION 5', mRNA sequence.
ACCESSION AA456707
VERSION AA456707.1 GI:2179283
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1877745.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m3 rev2 Er from Amersham
High quality sequence stop: 390.
Location/Qualifiers
FEATURES
source
1..404
/organism="Homo sapiens"
/db_xref="GDB:6044084"
/db_xref="taxon:9606"
/clone_lib="IMAGE:813174"
/tissue_type="pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three

normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT 71 a 153 c 114 g 66 t

ORIGIN

alignment_scores:
Quality: 551.00 Length: 111
Ratio: 5.009 Gaps: 0
Percent Similarity: 99.099 Percent Identity: 98.198

alignment_block:

US-09-135-238B-2_COPY_273_390 x AA456707 ..

Align seg 1/1 to: AA456707 from: 1 to: 404

8 LysAlaLeuSerArgArgAlaArgArgLeuAlaValArgMetArgAlaLe 24

:::|||||
1 GAAGCGCTCTCCAGCGCGCGCGCGACTGCGCGTGAGGATGCGCGCCT 50

24 uGluSerGlnArgProArgGlySerProArgProArgSerGlnAsnA 41

|||||
51 GGAGAGCTCCAGAGGCGCGCGGTCGCCGA.CCGCGCTCCCAAAACA 99

41 snLleYrSerAlaCysProArgArgAlaArgGlyAlaAspAlaLagly 57

|||||
100 ACATCTACAGCGCTGCGCGCGCGCTCGTGAGCGGACGCTGCAGGC 149

58 ThrGlyGluAlaProValProGlyProGlyAlaProLeuProAlaPr 74

|||||
150 ACAGGGAGGCGCGCGTTCGCGCGCGGAGCGCGTTCGCCCGCGCC 199

74 oLeuGlnValSerGluSerProfrpLeuHisAlaProSerLeuLysThrs 91

|||||
200 GCTCAGGTCTGTGAATCTCCCTGGCTCCATGCCCATCTCTGAAGACCA 249

91 erCysGluTyYrValSerLeuTyYrHisGlnProAlaAlaMetMetGluAsp 107

|||||
250 GCTGTGAATACGTGAGCTCTACACACAGCGCTGCGCGCATGATGGAGGAC 299

108 SerAspSerAspAspTyYrIleAsnValProAla 118

|||||
300 AGTGATTCAGATGACTACATCAATGTTCCTGCC 332

seq_name: gb_est42:AA401870

seq_documentation_block:

LOCUS AW401870 410 bp mRNA EST 16-FEB-2000
DEFINITION UI-HF-BK0-aah-g-11-0-UI.r1 NIH_MGC_36 Homo sapiens CDNA clone
IMAGE:3053781 5', mRNA sequence.

ACCESSION AW401870

VERSION AW401870.1 GI:6920556

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 410)

AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT On Jun 22, 1998 this sequence version replaced gi:3246884.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Eco RI site shown at the beginning of the sequence.

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.


```

242 AGCTCCCGCTTCTCCGAGGTACTTGAGCTCTCTGGCCACACC 291
||||| ||||| ||||| ||||| ||||| |||||
86 roSerLeuThrsSerCysGluTyrValSerLeuTyrHisGlnProAla 102
||||| ||||| ||||| ||||| ||||| ||||| |||||
292 CATCTCTGAAGATGAGCTGTGAATACGTGAGCTGGGCTACCAAGCTGCT 341
::: ||||| ||||| ||||| ||||| ||||| |||||
103 AlaMetMetGluAspSerAspSerAspTyrIleAsnValPro 117
::: ||||| ||||| ||||| ||||| ||||| |||||
342 GTCAACCTGGAAGACCTGATTCAGATGATTACATCAATATTCCT 386
::: ||||| ||||| ||||| ||||| ||||| |||||

```

seq_name: gb_est4:AA290194

seq_documentation_block: 471 bp mRNA EST 14-APR-1997
 LOCUS AA290194
 DEFINITION vb34f04.r1 Soares mouse lymph node NBMLN Mus musculus cDNA clone
 IMAGE:750847 5', mRNA sequence.

ACCESSION AA290194

VERSION AA290194.1 GI:1936558

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

AUTHORS 1 (bases 1 to 471)
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,B.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT On Dec 18, 1996 this sequence version replaced gi:1734376.
 Contact: Marra M/Mouse EST Project

WASHU-HMI Mouse EST Project

Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:459831

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 392.

Location/Qualifiers

1..471

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_lib="IMAGE:750847"

/sex="male"

/tissue_type="lymph node"

/dev_stage="4 weeks"

/lab_host="DH10B"

/note="Organ: lymph node; Vector: pT73D-Pac (Pharmacia)
 with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 1st strand cDNA was primed with a Not I - oligo(dT) primer
 [5'

TGTTACCAATCTGAAGTGGGAGCGCGGATACATCTTTTTTTTTTTTTTTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. RNA
 provided by Dr. Bertrand Jordan. Library constructed and
 normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 86 a 164 c 125 g 96 t

ORIGIN

alignment_scores:
 Quality: 251.00 Length: 117
 Ratio: 2.885 Gaps: 5

Percent Similarity: 74.359 Percent Identity: 55.556

alignment_block:

US-09-135-238b-2_COPY_273_390 x AA290194 ..

Align seg 1/1 to: AA290194 from: 1 to: 471

8 LysAlaLeuSerArgArgAlaArgArgLeuAlaValArgMetArgAlaLe 24
 ::::: ||||| ||||| ||||| ||||| ||||| |||||

42 GAAGCCTCTCCAGACAGCTGGCGGCGCTAGCATGAGGAGCGAGCGCG 91
 ::::: ||||| ||||| ||||| ||||| ||||| |||||

24 uGluSerSerGln.Arg.ProArgGlySerPro.....ArgPr 36
 ::::: ||||| ||||| ||||| ||||| ||||| |||||

92 GGGGCTTCGGCGCGCTCCACACAGCGCGGATGCTCGCAGACGA 141
 ::::: ||||| ||||| ||||| ||||| ||||| |||||

36 oArgSerGlnAsnAsnIleTyrSerAlaCysProArgArgAlaArgGlyA 53
 ::::: ||||| ||||| ||||| ||||| ||||| |||||

142 GCCTCCAGACAAACAGCTTACAGCGCTGCCCCCGGC.GCACGGGAC 190
 ::::: ||||| ||||| ||||| ||||| ||||| |||||

53 laAspAlaAlaGlyThrGlyGluAlaPro...ValProGlyProGlyAla 68
 ::::: ||||| ||||| ||||| ||||| ||||| |||||

191 CAGACAGCTTGGTCCAGCGGAGCTCCGCTCAACGCCCGCAGCTCA 240
 ::::: ||||| ||||| ||||| ||||| ||||| |||||

69 ProLeuProProAlaProLeuGlnValSerGluSerProTrpLeuHisAl 85
 ::::: ||||| ||||| ||||| ||||| ||||| |||||

241 CGCTCCCGCTTCGCCGA....GTACTGAGAGCTCTTGGCCCCACAC 286
 ::::: ||||| ||||| ||||| ||||| ||||| |||||

85 aProSerLeuLysThrSerCysGluTyrValSerLeuTyrHisGlnProA 102
 ::::: ||||| ||||| ||||| ||||| ||||| |||||

287 CCCATCTCTGAAGATGAGCTGTGAATACGTGGCTGAGCTGAGCTG 336
 ::::: ||||| ||||| ||||| ||||| ||||| |||||

102 laAlaMetMetGluAspSerAspSerAspTyrIleAsnValPro 117
 ::::: ||||| ||||| ||||| ||||| ||||| |||||

337 CTGTCAACCTGGAAGACCTTGATTCAGATGATTACATCAATATTCCT 383
 ::::: ||||| ||||| ||||| ||||| ||||| |||||

seq_name: gb_est21:AT468023

seq_documentation_block: 450 bp mRNA EST 14-APR-1999
 LOCUS AT468023

DEFINITION tJ84h03.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 IMAGE:2148245 3', mRNA sequence.

ACCESSION AT468023

VERSION AT468023.1 GI:4330113

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 450)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On Oct 8, 1998 this sequence version replaced gi:3727110.
 Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1729 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 444.

Location/Qualifiers

1..450

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"

/lab_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from five normalized
 libraries were mixed, and ss circles were made in vitro.

Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHSF pool 1: 309384-310919, 323208-325895 Soares NB2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HF8-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 123 a 97 c 99 g 131 t
ORIGIN

alignment_scores:
Quality: 210.00 Length: 38
Ratio: 5.526 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-135-238B-2_COPY_273_390 x AI468023 ..
Align seg 1/1 to: AI468023 from: 1 to: 450

81 ProTrpLeuHisAlaProSerLeuLysThrSerCysGluTyrValSerLe 97
|||||
321 CCTGGCTCCATGCCCATCTCTGAAGACGAGCTGTGAATACGTGAGCCT 370
97 uTyrHisGlnProAlaMetGluAspSerAspSerAspTyrI 114
|||||
371 CTACCACCGCTCCGCCCATGATGGAGGACAGTGCATTCAGATGACTACA 420
114 leAsnValProAla 118
|||||
421 TCAATGTTCTGCTGCC 434
seq_name: gb_est24:AI743020

seq_documentation_block:
LOCUS AI743020 521 bp mRNA EST 21-JUN-1999
DEFINITION wg85b10.x1 Soares NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:2371867 3' similar to TR:060667 060667 ANTI-FAS-INDUCED APOPTOSIS. ;, mRNA sequence.
ACCESSION AI743020
VERSION AI743020.1 GI:51111308
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Hominidae; Homo.
REFERENCE 1 (bases 1 to 521)
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Gibco
High quality sequence stop: 462.
Location/Qualifiers
1..521
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2371867"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI;

FEATURES
source

Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHSF pool 1: 309384-310919, 323208-325895 Soares NB2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HF8-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 137 a 122 c 115 g 147 t
ORIGIN

alignment_scores:
Quality: 210.00 Length: 38
Ratio: 5.526 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-135-238B-2_COPY_273_390 x AI743020 ..
Align seg 1/1 to: AI743020 from: 1 to: 521

81 ProTrpLeuHisAlaProSerLeuLysThrSerCysGluTyrValSerLe 97
|||||
319 CCTGGCTCCATGCCCATCTCTGAAGACGAGCTGTGAATACGTGAGCCT 368
97 uTyrHisGlnProAlaMetGluAspSerAspSerAspTyrI 114
|||||
369 CTACCACCGCTCCGCCCATGATGGAGGACAGTGCATTCAGATGACTACA 418
114 leAsnValProAla 118
|||||
419 TCAATGTTCTGCTGCC 432
seq_name: gb_est16:AI137500

seq_documentation_block:
LOCUS AI137500 398 bp mRNA EST 05-JUL-1999
DEFINITION UI-R-C2p-oc-a-09-0-UI-s1 UI-R-C2p Rattus norvegicus cDNA clone UI-R-C2p-oc-a-09-0-UI 3', mRNA sequence.
ACCESSION AI137500
VERSION AI137500.1 GI:3638277
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 398)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Apr 3, 1998 this sequence version replaced gi:3018677.
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.wesg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult spleen library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E.

```

seq_documentation_block:
LOCUS      AA174968      460 bp      mRNA      EST
DEFINITION ms78n03.r1 Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:617717
            5', mRNA sequence.
ACCESSION  AA174968
VERSION    AA174968.1 GI:1756098
            16-FEB-1997

```

4

SPROArgArgAla


```

|||||.....:::   :::   ::::   |||||
506 GCCCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 555
50 .....AlaArg.....
      |||||
556 CCGGCACACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGNGC 605
52 .....GlyAlaSpAlaAalaglyThrGlyG 60
      |||:|||||:|||||:|||||
606 GGTCCGTCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 655
60 luAlaProValProGlyProGlyAlaProLeuProAlaProLeuGln 76
      :|||:|||||:|||||:|||||:|||||
656 ACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAC..... 699
77 valSerGluSerProTrpLeuHisAlaPro 86
      |||:|||||
700 .....CACGCGCC 708

seq_name: gb_est46:AW731177

seq_documentation_block:
LOCUS AW731177 1210 bp mRNA EST 20-APR-2000
DEFINITION GA_Ea0010G17 Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Ea0010G17, mRNA sequence.
ACCESSION AW731177
VERSION AW731177.1 GI:7628835
KEYWORDS EST.
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Malvales;
Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 1210)
AUTHORS Leslie,A., Frisch,D., Yu,Y., Wood,T.C., Wing,R.A. and Wilkins,T.A.
TITLE An integrated analysis of the genetics, development, and evolution
of the cotton fiber
JOURNAL Unpublished (2000)
COMMENT On Jan 6, 2000 this sequence version replaced gi:6676833.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
High quality sequence stop: 1210.
FEATURES             source
     source
       1..1210
         /organism="Gossypium arboreum"
         /strain="AKA"
         /cultivar="8400"
         /db_xref="taxon:29729"
         /clone="GA_Ea0010G17"
         /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
         /tissue_type="Fibers isolated from bolls harvested 7-10
         dpa"
         /lab_host="E. coli"
         /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT    76 a    637 c    382 g    104 t    11 others
ORIGIN

alignment_scores:
Quality: 102.00 Length: 113
Ratio: 1.821 Gaps: 4
Percent Similarity: 49.558 Percent Identity: 31.858

alignment_block:
US-09-135-238B-2_COPY_273_390 x AW731177 ..
Align seq 1/1 to: AW731177 from: 1 to: 1210
```


OM of: US-09-135-238B-2_COPY_18_272 to: N_Geneseq_36.* out_format : pfs

Date: Sep 12, 2000 6:56 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/USPTO_pool/US09135238/runat_24082000_091324_17918/app_query.fasta_1.1282
-DB=N_Geneseq_36 -QFM=fastap -SUFFIX=p2n.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -GAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
-TRANS=human40.cdi -LIST=45 -DLOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTPMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=1000000
-USER=US09135238_cgn1_1_123 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-135-238B-2_COPY_18_272
Query length: 255
Database: N_Geneseq_36.*
Database sequences: 311585
Database length: 125096042
Search time (sec): 121.040000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
N_Geneseq_36:X28178	1340.00	2077.38	4.1e-108	2040	Human PIGRL-1 coding sequence.
N_Geneseq_36:X28179	1055.11	1605.11	8.4e-82	1047	Human PIGRL-1 coding sequence.
N_Geneseq_36:T31290	187.00	284.24	3.1e-08	3095	Mouse poly-immunoglobulin rec
N_Geneseq_36:T31422	188.00	247.67	3.4e-06	29392	Mouse poly Ig receptor prot
N_Geneseq_36:T31291	185.50	265.70	3.4e-07	3269	Rat poly-immunoglobulin rec
N_Geneseq_36:X07407	180.50	263.62	4.4e-07	1839	Rat secretory immunoglobulin
N_Geneseq_36:T30857	180.50	263.11	4.7e-07	1936	Secretory component DNA fragm
N_Geneseq_36:T30858	180.50	262.63	5.0e-07	2031	Secretory component DNA fragm
N_Geneseq_36:T31288	180.50	259.00	8.0e-07	2919	Human poly-immunoglobulin rec
N_Geneseq_36:T31289	175.50	255.26	1.3e-06	3630	Bovine poly-immunoglobulin rec
N_Geneseq_36:T31283	179.50	255.41	1.3e-06	1911	CDNA for human immunity relat
N_Geneseq_36:Q14498	166.50	241.51	7.5e-06	1876	Clone 1 for truncated poly Ig
N_Geneseq_36:T31287	166.50	235.23	1.7e-05	3517	Rabbit poly-immunoglobulin rec
N_Geneseq_36:Q14499	145.00	207.86	0.0006	1876	Clone 2 for truncated poly Ig
N_Geneseq_36:X28450	122.00	176.86	0.0299	1137	Human PIGR-2 coding sequence.
N_Geneseq_36:T96033	118.50	165.33	0.1313	2084	Rat kidney injury related mole
N_Geneseq_36:T96032	118.50	163.25	0.1714	2566	Rat kidney injury related mole
N_Geneseq_36:T96035	117.50	165.26	0.1325	1795	Human kidney injury related m
N_Geneseq_36:X87689	116.00	174.46	0.0407	565	EST clone EN260. New polynucle
N_Geneseq_36:X87689	116.00	174.46	0.0407	565	EST clone EN260. New polynucle
N_Geneseq_36:X34234	107.50	154.73	0.5115	1075	Human secreted protein gene. N
N_Geneseq_36:X34235	107.50	152.89	0.6477	1105	Human secreted protein gene. N
N_Geneseq_36:X34169	104.50	129.15	13.59	8690	Plasmid pD17-hm14.H1. Inhibit
N_Geneseq_36:V41432	104.00	146.42	1.48	1427	Plasmid Hu19BHCpcd encoding an
N_Geneseq_36:V41432	103.00	144.85	1.81	1427	Plasmid Hu19BHCpcd encoding an
N_Geneseq_36:V41431	103.00	144.85	1.81	1427	Plasmid Hu19BHCpcd encoding an
N_Geneseq_36:V18598	102.50	126.46	19.21	8321	Plasmid expressing hBR96-A. In
N_Geneseq_36:T15932	102.00	128.05	15.65	6557	Anti-IgE VH expression vector.
N_Geneseq_36:Q38746	101.50	134.47	6.87	3390	CD22-beta gene. Inhibition of
N_Geneseq_36:T62510	101.00	141.69	2.72	1431	Primate anti-human B7.1 ant
N_Geneseq_36:T35485	101.00	141.69	2.72	1431	Macaque primatized 7C10 heavy
N_Geneseq_36:Q35099	100.50	139.69	3.52	1617	Antibody D heavy chain. Prodn.
N_Geneseq_36:V41427	100.00	125.35	22.42	6284	Plasmid Hu19AHpcd encoding an
N_Geneseq_36:T60665	99.50	128.69	14.14	4157	Human CD10 antigen cDNA. Nucle
N_Geneseq_36:X07474	99.50	124.82	23.70	6127	Mus musculus anti-IgE antibody
N_Geneseq_36:X24074	99.00	143.04	2.29	944	Human EST R2810 DNA. Complexes
N_Geneseq_36:V66631	99.00	135.90	5.72	1869	Human immunoglobulin heavy cha
N_Geneseq_36:Q80470	98.50	143.79	2.08	784	Anti-phenytoxazone antibody c
N_Geneseq_36:V80580	98.50	140.78	3.06	1060	Kidney injury associated molec
N_Geneseq_36:Q49834	97.00	135.75	5.83	1386	Anti-HIV-1 recombinant antibod

N_Geneseq_36:Q12632 + 97.00 130.42 11.55 2364 ! CD4-specific CDR-grafted he
N_Geneseq_36:Q72708 - 97.00 110.51 148.55 17350 ! hOP1 human osteogenic prot
N_Geneseq_36:Q53142 - 97.00 110.47 149.20 17410 ! Sequence encoding osteogen
N_Geneseq_36:T18381 - 97.00 110.47 149.20 17410 ! hOP-1 genomic DNA. Antibod
N_Geneseq_36:V15205 - 97.00 110.47 149.20 17410 ! Human osteogenic protein O

seq_name: N_Geneseq_36:X28178

seq_documentation_block:

ID X28178 standard; cDNA; 2040 BP.
AC X28178:
DT 16-JUN-1999 (first entry)
DE Human PIGRL-1 coding sequence.
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;
KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;
KW diagnosis; therapy; ss.
OS Homo sapiens.
PN EP-905238-A2.
PD 31-MAR-1999. 306487.
PF 14-AUG-1998; US-961564.
PR 30-OCT-1997; US-961564.
PR 25-AUG-1997; US-056935.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Sweet RW, Truneh A, Wu S;
DR WPI: 99-192666/17.
DR P-PSDB; X05001.
PT New polypeptides encoding PIGRL-1 useful for treating diseases such
as X-linked Severe Combined Immunodeficiency
PS Claim 4; Page 18-19; 26pp; English.
CC This sequence encodes the human PIGRL-1 protein of the invention.
CC Autoimmune diseases involving altered expression or activity of PIGRL-1
may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined
Immunodeficiency (XSCID) and IgA deficiency. These diseases can be
diagnosed or susceptibility to them predicted by: (1) determining whether
there is a mutation in the genomic copy of the gene encoding PIGRL-1; or
(2) measuring the amount of PIGRL-1 in a sample derived from the patient.
Patients deficient in PIGRL-1 can be treated by administering either the
PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.
Patients with excessive expression or activity of PIGRL-1 can be treated
by administering an antagonist of PIGRL-1, an antisense nucleic acid
molecule which inhibits the expression of PIGRL-1 or administering
sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can
be used to identify its agonists by contacting a cell expressing PIGRL-1
with a candidate compound in the presence of a signal system and noting
the candidate as an agonist if a signal is produced. The same method can
be used to identify antagonists of PIGRL-1 but the presence of an
antagonist is indicated by a decrease in production of the signal.
Antibodies against PIGRL-1 may be used to isolate or identify clones
expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to
identify chromosomal mutations in the gene encoding PIGRL-1 in patients.
This information may then be correlated with the incidence of autoimmune
disease in those patients to identify whether the mutation causes the
disease.

alignment_scores:
Quality: 1340.00 Length: 255
Ratio: 5.255 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-135-238B-2_COPY_18_272 x X28178
Align seg 1/1 to: X28178 from: 1 to: 2040

1 ArgLeuProGluValLysValGluGlyGlySerValrth 17
|||||
197 AGGATCTCCAGAGTAAGTAGAGGGGAGCTGGCGGATCAGTTAC 246
|||||
17 rLeLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 34
|||||
247 CATCAAGTGCCCACTCTCTGAAATGCATGTGAGGATATATCTGTGCGGG 296

```

34  lMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
|||||
297 AGATGGCTGGATCTGGAACATGTGTACCGTGGTATCCACCAACTTC 346
|||||
51  IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67
|||||
347 ATCAGGCGAATACAAAGCGCGAGTACTCTGAGCAATACCCAGCAA 396
|||||
67  sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84
|||||
397 GAATCTGTTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGAG 446
|||||
84  alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
|||||
447 TCTATGCCCTGGAGCGGCGATGACACAGACCGGGGAAAGACCCAGAAA 496
|||||
101 ValThrLeuAsnValHisSerGlyTyrGluProSerTrpGluGlnPr 117
|||||
497 GTCACCCCTGAATGTCACAGTGAATACGAGCCATCATGGGAAGCAGCC 546
|||||
117 oMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMetP 134
|||||
547 AATGGCTGAGACTCAAAATGGTTTCATCTGCCCTATTGTTCCAGATGC 596
|||||
134 roAlaTyrAlaSerSerSerLysPheValThrArgValThrProAla 150
|||||
597 CTGCATATGCCAGTTCTTCCAAATTCGTAACACAGATTACACACAGCT 646
|||||
151 GlnArgGlyLysValProProValHisHisSerSerProThrThrGlnI 167
|||||
647 CAAAGGGGCAAGGTCCTCCAGTTCACCACTCCTCCCCACCACCCCAAT 696
|||||
167 eThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLysP 184
|||||
697 CACCCACCCGCCCTCAGTGTCCAGAGCATCTTCAGTAGCAGGTGACAAG 746
|||||
184 roArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuGlu 200
|||||
747 CCCGAACCTTCCTGCCATCCACTACAGCCTCAAAATCTCAGCTCTGGAG 796
|||||
201 GlyLeuLysProGlnThrProSerTyrAsnHisHisThrArgLeuHi 217
|||||
797 GGGCTGCTCAAGCCCCAGACCCAGCTACACACACACACACAGGCTGCA 846
|||||
217 sArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGlyGln 234
|||||
847 CAGCAGAGACACTGGACTATGGCTCACAGTCTGGGAGGAGGCCAAG 896
|||||
234 lPheHisIleLeuIleProThrIleLeuGlyLeuPheLeuAlaLeu 250
|||||
897 GATTTCACATCTGATCCCGCACCATCTGGGCCCTTTCTCTGTCGCACTT 946
|||||
251 LeuGlyLeuValVal 255
|||||
947 CTGGGGCTGGTGGTG 961
|||||
seq_name: N_Geneseq_36:X28179
seq_documentation_block:
ID X28179 standard; cDNA; 1047 BP.
AC X28179;
DT 16-JUN-1999 (first entry)
DE Human PIGRL-1 coding sequence fragment.
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;
KW x-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;
KW diagnosis; therapy; ss.
OS Homo sapiens.
PN EP-905238-A2.
PD 31-MAR-1999.
PF 14-AUG-1998; 306487.
PR 30-OCT-1997; US-961564.
PR 25-AUG-1997; US-056935.

```

(SMIK) SMITHKLINE BEECHAM CORP.
Sweet RW, Truneh A, Wu S;
WPI: 99-192666/17.
P-PSDB: Y05002.
New polypeptides encoding PIGRL-1 useful for treating diseases such as X-linked Severe Combined Immunodeficiency
Disclosure: Page 8; 26pp; English.
This sequence encodes the human PIGRL-1 protein of the invention.
Autoimmune diseases involving altered expression or activity of PIGRL-1 may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined Immunodeficiency (XSCID) and IgA deficiency. These diseases can be diagnosed or susceptibility to them predicted by: (1) determining whether there is a mutation in the genomic copy of the gene encoding PIGRL-1; or (2) measuring the amount of PIGRL-1 in a sample derived from the patient. Patients deficient in PIGRL-1 can be treated by administering either the PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient. Patients with excessive expression or activity of PIGRL-1 can be treated by administering an antagonist of PIGRL-1, an antisense nucleic acid molecule which inhibits the expression of PIGRL-1 or administering sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can be used to identify its agonists by contacting a cell expressing PIGRL-1 with a candidate compound in the presence of a signal system and noting the candidate as an agonist if a signal is produced. The same method can be used to identify antagonists of PIGRL-1 but the presence of an antagonist is indicated by a decrease in production of the signal. Antibodies against PIGRL-1 may be used to isolate or identify clones expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to identify chromosomal mutations in the gene encoding PIGRL-1 in patients. This information may then be correlated with the incidence of autoimmune disease in those patients to identify whether the mutation causes the disease.

Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T;

alignment_scores:
Quality: 1034.00 Length: 258
Ratio: 4.345 Gaps: 12
Percent Similarity: 92.248 Percent Identity: 89.922

alignment_block:
US-09-135-238b-2_COPY_18_272 x X28179 ..

Align seg 1/1 to: X28179 from: 1 to: 1047

1 ArgIleLeuProGluValLysValGluGlyLeuGlyGlySerValTh 17
|||||
201 AGGATCTCCACAGAACTAAAGGTACAGGGGAGCTGGCGGATCAGTTAC 250
|||||
17 rIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 34
|||||
251 CATCAAGTCCCACTTCTCTGAAATGCATGTGAGGATATATCTGTCGGGG 300
|||||
34 lMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
|||||
301 AGATGGCTGGATCTGGAACTATGTGTACCGTGGTATCCACCAACTTC 350
|||||
51 IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67
|||||
351 ATCAAGGCGAATACAAAGGCGCGAGTTACTCTGAAGCAATACCCAGCAA 400
|||||
67 sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84
|||||
401 GAATCTGTTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGAG 450
|||||
84 alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
|||||
451 TCTATGCCCTGGAGCGGCGATGACACAGACCGGGGAAAGACCCAGAAA 499
|||||
101 ValThrLeuAsnValHisSerGlyTyrGluProSerTrpGluGlnPr 117
|||||
500 GTCACCCCTGAATGTCACAGTGAATACGAGCCATCATGGGAAGCAGCC 549
|||||
117 oMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMetP 134

```

|||||
550 AATGCTCAGAGCTCAAAATGGTTTCATCTGCCCTATTTGTCAGATCG 599
134 toAlaTyrAlaSerSerSerLysPheValThr..ArgValThrThrProA 150
|||||
600 CTGCATATGCCGGTCTTCCACATTCGTAACCGCAGAGATTACCACACGAG 649
150 la.GlnArgGlyLysValProValHisHisSerSerProThrThrG1 166
|||||
650 CTTCAAGGGGCAAGGTCCTCCAGTTCACCACTCTCCCTCCCCACCCCA 699
166 nIle..ThrHisArgPro.ArgValSerArgAlaSerSerValAlaGlyAs 182
|||||
700 AATCACCACCGCCGCTTCGAGTTCNCGAGCATCTTCAGTAGCAGGTGA 749
182 pIysProArg..ThrPheLeuProSerThrThrAlaSerLysIleSerAla 198
|||||
750 CAAGCCCCGAAACITTCCTGCCATCCACTACAGGCTCAAAAATCTCAGCT 799
199 LeuGluGlyLeuLeuLys...ProGlnThrProSerTyrAsnHisHisTh 214
|||||
800 CTGGAAGGGCTGCTTCAAGCCCCAGAGCGCCAGCTACAAACAN.CACAC 848
214 tArgLeuHisArgGlnArgAlaLeuAsp...Tyr.GlySerGlnSerGly 229
849 CAGGCTGCACAGGAGAGAGACTGGATACTATTGGGNTCAGACTCTGGG 898
230 ArgGluGly.....GlnGlyPhe...HisIleLeuIlePro..ThrIleL 243
899 ...GAGGGGAANGNCCAGGATTTTACATTCCTGATTCCCGGACCATCN 945
243 euGlyLeuPhe 246
||| |||
946 FTGGGGGCTTT 956

```

seq_name: N_Geneseq_36:T31290

seq_documentation_block:

```

ID T31290 standard; cdNA; 3095 BP.
AC T31290;
DT 24-FEB-1997 (first entry)
DE Mouse poly-immunoglobulin receptor, cdNA.
KW Mouse; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; murine; ss.
OS Mus musculus.
PH Key Location/Qualifiers
FT cds 85..2400
FT /*tag= a
FN W09621012-A1.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PR (PLAN-) PLANT BIOTECHNOLOGY INC..
PR (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PR (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI; 96-333987/33.
DR P-PSDB: W03180.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure; pages 117-121; 152pp; English.
CC The present sequence encodes the mouse poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450.
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.

```

CC gastrointestinal, environments, therefore enhancing its
 CC effectiveness in passively immunising animals against mucosal
 CC pathogens. The Ag binding domain is specifically derived from the
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
 CC S. sorbinus serotypes d and g.
 SQ Sequence 3095 BP; 861 A; 796 C; 784 G; 654 T;

alignment_scores:
 Quality: 197.00 Length: 148
 Ratio: 2.165 Gaps: 5
 Percent Similarity: 61.486 Percent Identity: 32.432

alignment_block:

US-09-135-238B-2_copy_18_272 x T31290 ..

Align seg 1/1 to: T31290 from: 1 to: 3095

```

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
||| |||||
181 GGCACACTCTGTTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 230
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
|||:|||| ||| |||||
231 GCACACCCGGGAATACTGTTGCCGACACAGGAGCC...AGCGGCATGTGCA 277
42 lyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 58
|||||
278 CAAGGCTCATCTCTTCAATGGGTACCTCTCCAAGGAGTATTTCAGGCAGA 327
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||||
328 GCCAACCTCATCAACITCCAGAGAACACACATTTGTGATTACCATGA 377
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92
|||||
378 GCAGCTCACCCAGGACGACACTGGGAGCTACAAGTGTGGCTGGGT...A 424
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
425 CCAGTACCCGAGGCTGTCTTCGATGTGAGCTGGAGTGC..... 465
109 TyrGluProSerTyrGluGlnProMetProGluThrProLysTrpPh 125
|||
466 .....AGCCAGGTTCTCTGAGTTGCCGAGTGACAC 494
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 142
|||||
495 CCAGGTC.....TACACAAGGACATAGGCA 520
142 heValThrArgValThrThrProAlaGlnArgGlyLysValPro 156
|||
521 GAAATGTGACCATTTGAATGCCITTCAAAAGGAGGAATGTTCC 564

```

seq_name: N_Geneseq_36:V15422

seq_documentation_block:

```

ID V15422 standard; DNA; 29392 BP.
AC V15422;
DT 11-JUN-1998 (first entry)
DE Mouse poly Ig receptor protein gene.
KW Mouse; poly Ig receptor protein; pIgR protein; pIg; deficiency;
KW knockout mouse; disease model; ds.
OS Mus sp.
PN J10057066-A.
PD 03-MAR-1998.
PF 19-AUG-1996; 217154.
PR 19-AUG-1996; JP-217154.
PR (HONS) YAKULT HONSHA KK.
DR WPI; 98-254323/23.
PT Mouse pIg receptor protein gene - used for preparing gene knockout
PT mice, useful for study of human poly Ig receptor protein deficiency

```

PS Claim 1; Page 4-14; 18pp; Japanese.

CC The present sequence represents the mouse poly Ig receptor protein gene, which has a 29392 bp sequence. The new gene can be used to produce a gene knockout mouse, useful as a disease model of human poly Ig receptor protein deficiency.

SQ Sequence 29392 BP; 8318 A; 6747 C; 6514 G; 7813 T;

alignment_scores:

Quality: 188.00 Length: 97
Ratio: 2.648 Gaps: 3
Percent Similarity: 73.196 Percent Identity: 40.206

alignment_block:

US-09-135-238B-2_COPY_18_272 x V15422 ..

Align seg 1/1 to: V15422 from: 1 to: 29392

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
||||| ||||| ||||| ||||| |||||
18245 GCGGACTCTGTTCCATCATCGTGTACTACCCAGACACCTCTGTCAACCG 18294

25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18295 GCACACCCGGAATACTGTGTCGCCAGAGAGCC...AGCGGATGTGCA 18341

42 LyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 58
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18342 CAACGCTCATCTTCAATGGCTACCTCTCCAGAGATATTTCAGGCAGA 18391

59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18392 GCCAAGCTCATCACTTCCAGAGACACACATTTGTGATTAAACATTGA 18441

75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18442 GCAGCTCACCCAGGAGCAGCTGGGAGCTACAAGTGTGGCTGGGT...A 18488

92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 105
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18489 CAGATACCCAGGAGCCCTGCTTCGATGTCAGCCTGGAGGTC 18529

seq_name: N_Geneseq_36:T31291

seq_documentation_block:

ID T31291 standard; CDNA; 3269 BP.
AC T31291; 1997 (first entry)
DT 24-FEB-1997 (first entry)
DE Rat poly-immunoglobulin receptor, cDNA.
KW Rat; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; ss.
OS Rattus rattus.
FH Key Location/Qualifiers
FT cds 74..2383 /*tag= a
FT WO9621012-A1.
PN 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNWE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI; 96-333987/33.
DR p-PSDB; W03181.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure; Pages 123-127; 152pp; English.

CC The present sequence encodes the rat poly-immunoglobulin (Ig) receptor, a portion of which corresp. to residues 1-627, pref. CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450, CC 440-550, 550-606 or 550-627 comprises a protection protein (pp). CC The Ig of the invention comprises a pp as above in association with CC an Ig derived heavy chain, having at least a portion of an antigen CC (Ag) binding domain. The pp protects the Ig in harsh mucosal, e.g. CC gastrointestinal, environments, therefore enhancing its CC effectiveness in passively immunising animals against mucosal CC pathogens. The Ag binding domain is specifically derived from the CC Guy's 13 antibody, and the Ig can be used to prevent dental caries CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or CC S. sorbinus serotypes d and g.
SQ Sequence 3269 BP; 890 A; 831 C; 850 G; 698 T;

alignment_scores:

Quality: 185.50 Length: 164
Ratio: 1.855 Gaps: 6
Percent Similarity: 60.976 Percent Identity: 30.488

alignment_block:

US-09-135-238B-2_COPY_18_272 x T31291 ..

Align seg 1/1 to: T31291 from: 1 to: 3269

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
170 GGTAACTCGTCTCCATCATCGTGTACTACCCAGACACCTCTGTCAACCG 219

25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
220 GCACACCCGGAATACTGTGTCGCCAGAGAGCC...AACGGTACTGGG 266

42 LyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 58
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
267 CAACCTCATCTTCAATGGCTACCTCTCGAAGAGATATTTCAGGCAGA 316

59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 GCAGCTCATCACTTCCAGAGACACATTTGTGATTAAACATTTC 366

75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
367 ACATCTCACCCAGGAGACACTGGGAGCTACAAGTGTGGTCTGGGT...A 413

92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
414 CCATTAACCCAGAGCCCTGTTTTCGATGTCAGCCTGGAGGTC..... 454

109 TyrGluProSerTrpGluGlnProMetProGluThrProLysTrpPh 125
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
455AGCCAGGTTCTCTGAGTCCCAAAATGACAC 483

125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 142
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
484 CCATGTC.....TACACAAAGG 500

142 heValThrArgValThrThrProAlaGlnArgGlyLysValProProVal 158
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
501 ACATAGGCAGAACTGTGACCATCGAATGCCGTTTCAAGAGGGGAATGCT 550

159 HisHisSer_SerProThrThrGlnIleThrHisArgPro 171
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
551 CATAGCAAGAAATCCCTGTGTAAGAGAGAGAGAGGCGCT 590

seq_name: N_Geneseq_36:X07407

seq_documentation_block:

ID X07407 standard; CDNA to mRNA; 1839 BP.
AC X07407;
DT 08-JUN-1999 (first entry)
DE Human secretory immunoglobulin A component gene.

Immunoglobulin A; secretory; component; IgA; human; treatment; prevention; infection; HIV; AIDS; cold; flu; virus; gene; human immunodeficiency virus; respiratory syncytial virus; ss. Homo sapiens.

Key Location/Qualifiers

CDS 1..1839

FT /*tag= a

FT /product= secretory component

PN WO9857993-A1.

PD 23-DEC-1998.

PF 10-JUN-1998; U11975.

PR 19-JUN-1997; US-050969.

PA (REGC) UNIV CALIFORNIA.

PI Chintalacharuvu KR, Morrison SL;

DR WPI; 99-080950/07.

DR P-PSDB; W95601.

PT Producing secretory immunoglobulin in single cells - useful to produce commercial quantities of secretory immunoglobulin to prevent or treat infections

PT Disclosure: Page 22; 39pp; English.

CC The sequence is that of the coding region for the secretory component of human secretory immunoglobulin A (sIgA).

CC The sequence can be used as part of a method for the production of sig molecules. This method is useful for producing commercial quantities of sig (especially sIgA) to treat or prevent infections. In particular, sIgA produced by the method can be used to prevent or treat infections in mammals, birds or fish; especially systemic infections or infections at a mucosal surface. It is especially useful to prevent or treat infection with human immunodeficiency virus (HIV), respiratory syncytial virus, flu virus or cold virus. The method allows production of commercial quantities of sig molecules for therapeutic use, not previously possible; production using non-plant cells and a single cell type is more efficient than a previous multi-step process of fusing recombinant plant cells, and avoids alterations of the sig by plant cells. SigA molecules are more stable and resistant to proteolysis than previously used IgA molecules, and can be administered to prevent as well as to treat infections, unlike e.g. IgG and IgM molecules.

SQ Sequence 1839 BP; 423 A; 504 C; 544 G; 368 T;

alignment_scores:

Quality: 180.50 Length: 106

Ratio: 2.314 Gaps: 4

Percent Similarity: 73.585 Percent Identity: 38.679

alignment_block:

US-09-135-238B-2_COPY_18_272 x X07407 ..

Align seg 1/1 to: X07407 from: 1 to: 1839

4 ProGluValLysValGluGlyGluLeuGlyGlySerValThrIleLysCys 20

73 CCCGAG...GAGGTGAATAGTGTGGAAGGTAACCTCAGTGTCCATCAGTG 119

20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33

120 CTACTACCCACCCACCTCTGTCAACGGGCACACCCGGAGTACTGTGTGCC 169

33 rGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49

170 GGCAGGGAGCTAGAGGTGC...TGCATAACCTCATCTCTCGGAGGCG 216

50 pheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66

217 TACGTCTCCAGCAATATGCGAGCGAGGGTAACTCACCACCACTCCCGGA 266

66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83

267 GAACGGCACATTGTGGTGRACATGTCACAGCTGACCGAGGATGACTCCG 316

83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99

317 GGCGCTACAAGTGTGGCTGGGCATCAATAGC...CGAGGCTGTCTCTTT 363

100 LysValThrLeuAsnVal 105

364 GATGTACGCTGGAGGTC 381

seq_name: N_Geneseq_36:T30857

seq_documentation_block:

ID T30857 standard; cDNA; 1936 BP.

AC T30857;

DE 13-SEP-1996 (first entry)

DE Secretory component DNA fragment 2.

KW Secretory component; SC DNA fragment 2; crystallisation; polymeric immunoglobulin receptor; pIgR; CHO; ss.

OS Chimeric Homo sapiens;

OS Chimeric synthetic.

FT Key Location/Qualifiers

FT misc_feature 1..46

FT /*tag= a

FT /note= "Bases 1-46 derived from PCR with 5' primer 2"

FT mat_peptide 16..1920

FT /*tag= b

FT /product= transcript from SC DNA fragment 2 complement (1902..1936)

FT /*tag= c

FT /note= "from PCR with 3' primer"

PN WO9618734-A1.

PD 20-JUN-1996.

PF 06-DEC-1995; EC4797.

PR 16-DEC-1994; EP-120019.

PA (CIBA) CIBA GEIGY AG.

PI Asselbergs FAM, Gruetter M, Leist C, Van Oostrum J;

DR WPI; 96-300651/30.

PT Prodn. of recombinant secretory protein capable of crystallisation - pref. human poly-IgR, for use in screening and binding studies

PS Claim 5; Page 39-41; 50pp; English.

CC SC DNA fragment 2 (T30857) was obt'd. by PCR amplification of human polymeric immunoglobulin receptor (pIgR) cDNA in vector pCB6 using a 5' primer (T30860) that creates an HindIII site upstream of the ATG start codon and a 3' primer (T30861) that generates a stop codon at position 1906 just upstream of the hydrophobic transmembrane segment of the pIgR and also creates a downstream XbaI site. CHO SSF 3 cells transfected with a vector carrying the amplified DNA can be used for prodn. of recombinant secretory component (SC). This can be crystallised for receptor structure studies, used to stabilise immunoglobulins, or used to screen (ant)agonists capable of modulating mucosal immune responses.

SQ Sequence 1936 BP; 452 A; 525 C; 578 G; 381 T;

alignment_scores:

Quality: 180.50 Length: 106

Ratio: 2.314 Gaps: 4

Percent Similarity: 73.585 Percent Identity: 38.679

alignment_block:

US-09-135-238B-2_COPY_18_272 x T30857 ..

Align seg 1/1 to: T30857 from: 1 to: 1936

4 ProGluValLysValGluGlyGluLeuGlyGlySerValThrIleLysCys 20

88 CCCGAG...GAGGTGAATAGTGTGGAAGGTAACCTCAGTGTCCATCAGTG 134

20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33

135 CTACTACCCACCCACCTCTGTCAACGGGCACACCCGGAGTACTGTGTGCC 184

33 rGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49

185 GCGAGGAGCTAGAGTGGC...TGCATACCCCTCATCTCTCGGAGGCG 231
 50 PhelelLysAlaCluTyLysGlyArgValThrLeuLysGlnTyProAr 66
 232 TACGTCTCCAGCAATATGCGAGGAGGCTAACCTTCCACCACTTCCCGGA 281
 66 gLyAsnLeuPheLeuValGluValThrGlnLeuThrGluSerG 83
 282 GAAGGACACATTCGTGTGAACATTGCCAGCTGAGCAGGATGACTCG 331
 83 LysValTyraLysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
 332 GCGCTACAGTGTGGCTGGCATCATAGC...CGAGGCTGTCTCTT 378
 100 LysValThrLeuAsnVal 105
 379 GATGTCAGCCTGGAGGTC 396

seq_name: N_Geneseq_36:T30856

seq_documentation_block:
 ID T30856 standard; cDNA; 2031 BP.
 AC T30856;
 DT 13-SEP-1996 (first entry)
 DE Secretary component DNA fragment 1.
 KW Secretory component; SC DNA fragment 1; crystallisation;
 KW polymeric immunoglobulin receptor; pIgr; CHO; ss.
 OS Chimeric Homo sapiens;
 OS Chimeric synthetic.
 FH Key Location/Qualifiers
 FT misc_feature 1..33
 FT /*tag= a
 FT /note= "bases 1-33 derived from PCR with 5'
 FT primer 1"
 FT 111..2015
 FT /*tag= b
 FT /product= transcript from SC DNA fragment 1
 FT complement (1997..2031)
 FT /*tag= c
 FT /note= "from PCR with 3' primer"
 FT WO9618734-AL.
 PN 20-JUN-1996.
 PD 06-DEC-1995; E04797.
 PF 16-DEC-1994; EP-120019.
 PR (CIBA) CIBA GEIGY AG.
 PA Asselbergs FAM, Gruetter M, Leist C, Van Oostrum J;
 PI WPI; 96-300651/30.
 DR Prodn. of recombinant secretory protein capable of crystallisation -
 PT pref. human poly-IgR, for use in screening and binding studies
 PS Claim 5; Page 36-38; 50pp; English.
 CC SC DNA fragment 1 (T30856) was obtd. by PCR amplification of human
 CC polymeric immunoglobulin receptor (pIgr) cDNA in vector pCB6 using a
 CC 5' primer (T30859) hybridising to the cytomegalovirus promoter in
 CC pCB6 and a 3' primer (T30861) that generates a stop codon at
 CC position 1906 just upstream of the hydrophobic transmembrane segment
 CC of the pIgr and also creates a downstream XbaI site. CHO SSF 3
 CC cells transfected with a vector carrying the amplified DNA can be
 CC used for prodn. of recombinant secretory component (SC). This can
 CC be crystallised for receptor structure studies; used to stabilise
 CC immunoglobulins, or used to screen (ant)agonists capable of
 CC modulating mucosal immune responses.
 SQ Sequence 2031 BP; 477 A; 541 C; 605 G; 408 T;

alignment_scores:
 Quality: 180.50 Length: 106
 Ratio: 2.314 Gaps: 4
 Percent Similarity: 73.585 Percent Identity: 38.679

alignment_block:
 US-09-135-238b-2_COPY_18_272 x T30856 ..
 Align seg 1/1 to: T30856 from: 1 to: 2031

4 ProGluValLysValGluGlyGluLeuGlyGlySerValThrIleLysCy 20
 183 CCGAG...GAGTGAATAGTGTGGAGGTAACCTAGTCTCCATCAGTG 229
 20 sProLeuPro.....GluMetHisValArgIleTyLeuCySA 33
 230 CTACTACCCACCCACTCTGTCAACCGGCACACCGGAGTACTGTGTC 279
 33 TgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
 280 GCGAGGAGCTAGAGTGGC...TGCATACCCCTCATCTCTCGGAGGCG 326
 50 PhelelLysAlaGlyTyLysGlyArgValThrLeuLysGlnTyProAr 66
 327 TACGTCTCCAGCAATATGCGAGGAGGCTAACCTTCCACCACTTCCCGGA 376
 66 gLyAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83
 377 GAAGGACACATTCGTGTGAACATTGCCAGCTGAGCAGGATGACTCG 426
 83 LysValTyraLysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
 427 GCGCTACAGTGTGGCTGGCATCATAGC...CGAGGCTGTCTCTT 473
 100 LysValThrLeuAsnVal 105
 474 GATGTCAGCCTGGAGGTC 491
 seq_name: N_Geneseq_36:T31288
 seq_documentation_block:
 ID T31288 standard; cDNA; 2919 BP.
 AC T31288;
 DT 24-FEB-1997 (first entry)
 DE Human poly-immunoglobulin receptor, cDNA.
 KW Human; immunoglobulin; receptor; protection protein; mutants;
 KW heavy chain; antigen binding domain; protection; pathogen;
 KW mucosal; environment; gastrointestinal; passive; immunisation;
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
 KW poly; sorbinus; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 235..2475
 FT /*tag= a
 FT WO9621012-AL.
 PN 11-JUL-1996.
 PD 27-DEC-1995; U16889.
 PR 30-DEC-1994; US-367395.
 PR 04-MAY-1995; US-434000.
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.
 PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
 PI (PLAN-) PLANET BIOTECHNOLOGY INC.
 PI Hiatt AC, Lehner T, Ma JMC;
 DR WPI; 96-333987/33.
 DR P-PSDB; W03178.
 PT Immunoglobulin and protection protein complex and its prodn. in
 PT plants - useful for passive immunisation against mucosal antigens,
 PT esp. against S. mutans and S. sorbinus to prevent dental caries
 PS Disclosure; Pages 105-108; 152pp; English.
 CC The present sequence encodes the human poly-immunoglobulin (Ig)
 CC receptor, a portion of which corresp. to residues 1-627, pref.
 CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
 CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
 CC The Ig of the invention comprises a PP as above in association with
 CC an Ig derived heavy chain, having at least a portion of an antigen
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
 CC gastrointestinal, environments, therefore enhancing its
 CC effectiveness in passively immunising animals against mucosal
 CC pathogens. The Ag binding domain is specifically derived from the
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
 CC S. sorbinus serotypes d and g.

```
SQ Sequence 2919 BP; 684 A; 806 C; 854 G; 575 T;

alignment_scores:
  Quality: 180.50 Length: 106
  Ratio: 2.314 Gaps: 4
  Percent Similarity: 73.585 Percent Identity: 38.679

alignment_block:
US-09-135-238B-2_COPY_18_272 x T31288 ..
Align seg 1/1 to: T31288 from: 1 to: 2919
4 ProGluValLysValGluGluLeuGlyGlySerValThrIleLysCys 20
||||| :|||:|||||:|||||:|||||:|||||:
253 CCCGAG...GAGGTGATAGTGTGGAAGGTAACCTCAGTGTCCATCAGTG 299
||||| :|||:|||||:|||||:|||||:|||||:
20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33
||||| :|||:|||||:|||||:|||||:|||||:
300 TACTATACCCACCCACCTCTGTCAACCGGCACACCCGGAAGTACTGTGTC 349
||||| :|||:|||||:|||||:|||||:|||||:
33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
||||| :|||:|||||:|||||:|||||:|||||:
350 GCGAGGAGCTAGAGGTGGC...TGCATAACCTCATCTCCTCGGAGGCG 396
||||| :|||:|||||:|||||:|||||:|||||:
50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66
||||| :|||:|||||:|||||:|||||:|||||:
397 TAGGTCTCCGCAATATGCGAGGCGGTAACTCACCACCACTTCCCGGA 446
||||| :|||:|||||:|||||:|||||:|||||:
56 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83
||||| :|||:|||||:|||||:|||||:|||||:
447 GAACGGCACATTTGTGTGACATGTCACCTGAGCCGNGGATGACTCCG 496
||||| :|||:|||||:|||||:|||||:|||||:
83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
||||| :|||:|||||:|||||:|||||:|||||:
497 GCGCTACAGTGTGGCTGGGCAATAGC...CGAGGCGCTGTCTCTT 543
||||| :|||:|||||:|||||:|||||:|||||:
100 LysValThrLeuAsnVal 105
||||| :|||:|||||:|||||:|||||:|||||:
544 GATGTGACGCTGGAGGTC 561

seq_name: N_Geneseq_36:T31289

seq_documentation_block:
ID T31289 standard; cDNA; 3630 BP.
AC T31289;
DT 24-FEB-1997 (first entry)
DE Bovine poly-immunoglobulin receptor, cDNA.
KW Bovine; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's l3 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; cow; ss.
OS Bos taurus.
FH Key Location/Qualifiers
FT cds 152..2425
FT /*tag= a
FT WO9621012-A1.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNWE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PI Haatt AC, Lehner T, Ma JKC;
DR WPI; 96-333987/33.
DR P-PSDB; W03179.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure; pages 111-115; 152pp; English.
CC The present sequence encodes the bovine poly-immunoglobulin (Ig)
```

```
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's l3 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 3630 BP; 880 A; 1018 C; 1029 G; 703 T;

alignment_scores:
  Quality: 179.50 Length: 109
  Ratio: 2.272 Gaps: 4
  Percent Similarity: 72.477 Percent Identity: 37.615

alignment_block:
US-09-135-238B-2_COPY_18_272 x T31289 ..
Align seg 1/1 to: T31289 from: 1 to: 3630
4 ProGluValLysValGluGluLeuGlyGlySerValThrIleLysCys 20
||||| :|||:|||||:|||||:|||||:|||||:
224 CCCGAG...GAGGTGACGCGGTGGAAGCGCGCTCAGTGTCCATCAAGTG 270
||||| :|||:|||||:|||||:|||||:|||||:
20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33
||||| :|||:|||||:|||||:|||||:|||||:
271 TCTATACCCGCCACCTCCGTCACCGGCACACGCGCAAGTACTGTGTCG 320
||||| :|||:|||||:|||||:|||||:|||||:
33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
||||| :|||:|||||:|||||:|||||:|||||:
321 GCGAGGAGGCG...CAGGCGCGCTGCACGACCTCATCTCTCGGAGGCG 367
||||| :|||:|||||:|||||:|||||:|||||:
50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66
||||| :|||:|||||:|||||:|||||:|||||:
368 TAGCTCTCCGACGACTACGTGGGCGAGAGCCACCTCACCACCACTTCCCGGA 417
||||| :|||:|||||:|||||:|||||:|||||:
66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83
||||| :|||:|||||:|||||:|||||:|||||:
418 GAGCGGCACGTTGTGTGGTACATCAGCCATCTCACCATAAAGACTCAG 467
||||| :|||:|||||:|||||:|||||:|||||:
83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
||||| :|||:|||||:|||||:|||||:|||||:
468 GCGCTACAAAGTGTGGCTGGCATTAGCAGC...CGTGGCCTTAACCTTC 514
||||| :|||:|||||:|||||:|||||:|||||:
100 LysValThrLeuAsnValHisSerGlu 108
||||| :|||:|||||:|||||:|||||:|||||:
515 GATGTGAGCGCTGGAGGTCAGCCCAAGAT 541

seq_name: N_Geneseq_36:V20383

seq_documentation_block:
ID V20383 standard; cDNA to mRNA; 1911 BP.
AC V20383;
DT 26-JUN-1998 (first entry)
DE cDNA for human immunity related factor.
KW Lymph node; human; immunity related factor; research; treatment;
KW immune disease; infectious disease; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 98..1696
FT /*tag= a
FT sig_peptide 98..145
FT /*tag= b
FT mat_peptide 146..1693
FT /*tag= c
FT /product= immunity_related_factor
PT J10072495-A.
PD 17-MAR-1998.
```


Align seg 1/1 to: Q14498 from: 1 to: 1876

```

13 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 25
   ||| |||||:||||| |||
238 GGGGACTCGGTGTCATCATCTACTACCCCAACACCTCGTCACCG 287
   ||| ||| :||| |||||
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
   ||| ||| :||| |||||
288 GCACAGCCGGAAGTTCTGTGCGGGAA...GAGGAGAGCGCCGCTGG 334
   ||| ||| :||| |||||
42 lyThrValValSerThrThrAsnPhelLysAlaGluTyrLysGlyArg 58
   ||| ||| :||| |||||
335 TGACGCTTCCTCG...ACGGGTACACGTCCAGGAATACTCCGGGAGA 381
   ||| ||| :||| |||||
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
   ||| ||| :||| |||||
382 GGCAAGCTCAGGACTTCCTGATAAAGGGGAGTTGTGTGACTGTTGA 431
   ||| ||| :||| |||||
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
   ||| ||| :||| |||||
432 CCAACTCACCCAGACGACTCAGGAGCTACAAGTGTGCGTGGGAGTCA 481
   ||| ||| :||| |||||
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
   ||| ||| :||| |||||
482 AC...GGCGTGGCTGGACTTCGTGTCAACGTGTCAGCCAGGAAG 528
   ||| ||| :||| |||||
109 TyrGluPro 111
   ||| |||
529 CCAGAGCCT 537

```

seq_name: N_Geneseq_36:T31287

seq_documentation_block:

```

AC T31287 standard; cDNA; 3517 BP.
DT 24-FEB-1997 (first entry)
DE Rabbit poly-immunoglobulin receptor, cDNA.
KW Rabbit; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; ss.
OS Oryctolagus cuniculus.
FH Key Location/Qualifiers
FT 124..245
FT cds /*tag= a
PN W09621012-A1.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PI Hiatt AC, Lehner T, Ma JKC.
DR WPI; 96-333987/33.
DR P-PSDB; W03177.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PT Claim 10; Pages 99-102; 152pp; English.
PS The present sequence encodes the rabbit poly-immunoglobulin (Ig)
PS receptor, a portion of which corres. to residues 1-627, pref.
CC 1-606, or esp. residues 21-43, 1-118, 119-223, 224-332, 333-441,
CC 442-552, 553-606 or 533-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries

```

```

CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 3517 BP; 795 A; 1027 C; 1038 G; 657 T;

```

alignment_scores:

```

Quality: 166.50 Length: 103
Ratio: 2.602 Gaps: 4
Percent Similarity: 62.136 Percent Identity: 42.718

```

alignment_block:

US-09-135-238B-2_COPY_18_272 x T31287 ..

Align seg 1/1 to: T31287 from: 1 to: 3517

```

13 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 25
   ||| |||||:||||| |||
238 GGGGACTCGGTGTCATCATCTACTACCCCAACACCTCGTCACCG 287
   ||| ||| :||| |||||
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
   ||| ||| :||| |||||
288 GCACAGCCGGAAGTTCTGTGCGGGAA...GAGGAGAGCGCCGCTGG 334
   ||| ||| :||| |||||
42 lyThrValValSerThrThrAsnPhelLysAlaGluTyrLysGlyArg 58
   ||| ||| :||| |||||
335 TGACGCTTCCTCG...ACGGGTACACGTCCAGGAATACTCCGGGAGA 381
   ||| ||| :||| |||||
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
   ||| ||| :||| |||||
382 GGCAAGCTCAGGACTTCCTGATAAAGGGGAGTTGTGTGACTGTTGA 431
   ||| ||| :||| |||||
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
   ||| ||| :||| |||||
432 CCAACTCACCCAGACGACTCAGGAGCTACAAGTGTGCGTGGGAGTCA 481
   ||| ||| :||| |||||
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
   ||| ||| :||| |||||
482 AC...GGCGTGGCTGGACTTCGTGTCAACGTGTCAGCCAGGAAG 528
   ||| ||| :||| |||||
109 TyrGluPro 111
   ||| |||
529 CCAGAGCCT 537

```

seq_name: N_Geneseq_36:Q14499

seq_documentation_block:

```

ID Q14499 standard; cDNA; 1876 BP.
AC Q14499;
DT 30-JAN-1992 (first entry)
DE Clone 2 for truncated poly Ig-receptor.
KW Rabbit; insemination; pregnancy; ss.
OS Oryctolagus cuniculus.
FH Key Location/Qualifiers
FT cds 124..1876
PN W09116061-A.
PD 31-OCT-1991.
PF 16-APR-1991; U02604.
PR 16-APR-1990; US-510161.
PA (HARD ) HARVARD COLLEGE.
PA (SURE-) INST SUISSE RECH EXPER C.
PI Kraehenbuhl JP, Weltzin RA, Neutra MR;
DR WPI; 91-339549/46.
DR P-PSDB; R14671.
PT Stabilised poly-Ig complex contg. portion of poly-Ig receptor -
PT useful in protection against pathogens or against pregnancy
PS Disclosure; Fig 8; 51 pp; English.
CC The sequence was obtd. from a cDNA clone of allele no. 2 and
CC encodes a truncated poly-Ig receptor. The native gene (Mostov et
CC al) is mutated to delete the portion encoding the transmembrane and
CC intra-cellular domains. The recombinant protein produced by
CC expression of the sequence is used as a stabiliser protein with a
CC poly-Ig specific for a selected antigen or family of antigens. The
CC compsn. can be administered directly to the mucosal surfaces of a

```


206	GlnThrProSerTyrAsnHisHisThrArgLeuHisArgClnArgAlaLeu	222
635	635
222	uAspTyrGlySerGlnSerGlyArgGluGlyGlnGlyPheHisIleLeuI	239
636AATTcAGGGTTCcGGGcTCAGcAGcCCcTCACcTTCCcTGC	672
239	leProThrIleLeuGlyLeu...PheLeuLeuAlaLeuLeuGlyLeuVal	254
673	TCGtGGTcCTcTTCtGAAGcTcGCCcTcGTcCTcTcAGcATcGTcGGTcGCTcTC	722

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2000, 10:09:19 ; Search time 3928.49 seconds
(without alignments)
2144.023 Million cell updates/sec

Title: US-09-135-238B-1
Perfect score: 1910
Sequence: 1 aaaggagtaagcagcggtgc.....ttactctgtctcctctt 1910

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues
Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1:	em_est1:*
2:	em_est2:*
3:	em_est3:*
4:	em_est4:*
5:	em_est5:*
6:	em_est6:*
7:	em_est7:*
8:	em_est8:*
9:	em_est9:*
10:	em_est10:*
11:	em_est11:*
12:	em_est12:*
13:	em_est13:*
14:	em_est14:*
15:	em_est15:*
16:	em_est16:*
17:	em_est17:*
18:	em_est18:*
19:	em_est19:*
20:	gb_est1:*
21:	gb_est2:*
22:	gb_est3:*
23:	gb_est4:*
24:	gb_est5:*
25:	gb_est6:*
26:	gb_est7:*
27:	gb_est8:*
28:	gb_est9:*
29:	gb_est10:*
30:	gb_est11:*
31:	gb_est12:*
32:	gb_est13:*
33:	gb_est14:*
34:	gb_est15:*
35:	gb_est16:*
36:	gb_est17:*
37:	gb_est18:*
38:	gb_est19:*
39:	gb_est20:*
40:	gb_est21:*
41:	gb_est22:*
42:	gb_est23:*
43:	gb_est24:*

44:	gb_est25:*
45:	em_est20:*
46:	em_est21:*
47:	em_est22:*
48:	em_est23:*
49:	em_est24:*
50:	em_est25:*
51:	em_est26:*
52:	em_est27:*
53:	em_est28:*
54:	em_est29:*
55:	em_est30:*
56:	em_est31:*
57:	em_est32:*
58:	em_est33:*
59:	em_est34:*
60:	em_est35:*
61:	em_est36:*
62:	em_est37:*
63:	gb_est26:*
64:	gb_est27:*
65:	gb_est28:*
66:	gb_est29:*
67:	gb_est30:*
68:	gb_est31:*
69:	gb_est32:*
70:	gb_est33:*
71:	gb_est34:*
72:	gb_est35:*
73:	gb_est36:*
74:	gb_est37:*
75:	gb_est38:*
76:	gb_est39:*
77:	gb_est40:*
78:	gb_est41:*
79:	gb_est42:*
80:	gb_est43:*
81:	gb_est44:*
82:	gb_est45:*
83:	gb_est46:*
84:	gb_est47:*
85:	gb_est48:*
86:	gb_est49:*
87:	gb_est50:*
88:	gb_est51:*
89:	gb_est52:*
90:	gb_est53:*
91:	gb_est54:*
92:	gb_est55:*
93:	gb_gss1:*
94:	gb_gss2:*
95:	gb_gss3:*
96:	gb_gss4:*
97:	em_gss1:*
98:	em_gss2:*
99:	em_gss3:*
100:	em_gss4:*
101:	gb_gss15:*
102:	gb_gss16:*
103:	gb_gss17:*
104:	gb_gss18:*
105:	gb_gss19:*
106:	em_gss5:*
107:	em_gss6:*
108:	em_gss7:*
109:	em_gss8:*
110:	em_gss9:*
111:	em_gss10:*
112:	em_gss11:*
113:	em_gss12:*
114:	em_gss13:*
115:	gb_gss5:*
116:	gb_gss6:*

COMMENT		On May 18, 1998 this sequence version replaced gi:3137192. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html Insert Length: 2024 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 453.	
FEATURES		Location/Qualifiers	
source		1..458	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone_image="2389821"	
		/clone_lib="NCI_CGAP_CL11"	
		/issue_type="B-cell, chronic lymphocytic leukemia"	
		/lab_host="DH10B"	
		/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCATGCTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT		126 a 105 c 108 g 118 t 1 others	
ORIGIN			
Query Match		20.9%; Score 400; DB 44; Length 458;	
Best Local Similarity		96.8%; Pred. No. 4e-99;	
Matches 429; Conservative		0; Mismatches 11; Indels 3; Gaps 2;	
QY	1417	ccaggtcgtcttgacacacttgccgccccctgttagtgacaggtgaagctgtaggcatgt	1476
Db	458	CCAGTTCGTCTTGACACCTTGGCAGCCCCCTGTAGTTGACAGTAAGCTGTAGGCATGT	399
QY	1477	agagcaattgtcccaatgcacactgtctcttcccaagccgtcgacagactgtggatt	1536
Db	398	AGAGCAATTGTCCCAATGCCACTTGTCTCTTCCAGCCGTCGAACAGACTGTGGGATT	339
QY	1537	tgcagagtgtttcttccatgtctttgaccacaggggtgtgtgtgtccaggctctagatc	1596
Db	338	TGCAGAGTGTCTTCCATGTCTTTGNACCAGGGGT-TGTTGTGCCAGGCTCTAGATC	280
QY	1597	acatggcatcaggctggggcagaggcatagotattgtctcgtggccatccctcccaagggtg	1656
Db	279	ACATGGCATCAGGCTGGGGCANAGGCATAGCTATTGTCTCGGCGCATCCTCCAGGGTTG	220
QY	1657	ggtttacacaaataaagagctcttactctgaatttatgaactgacctcagccatgga	1716
Db	219	GGTCTTTACAAATAAGAGGCTCTGTCTCTGAGTTATGTGACATGCTCAGGCCCNAGA	160
QY	1717	ctaagcagggtgtgtgtat--aaacactcctgtgaacccgttgcctgcatcaaatgtt	1774
Db	159	CTAAGCAGGGGTTTGGTATAAAAAACACTCTCTGAAAACGCCCTTTCGCTGATCCAAATGTT	100
QY	1775	agactgttagtgaaactctactatctcaagtctatgctaaaggcaattatctga	1834
Db	99	AGCACTTGTCTAGTGAGCGTCTACTTATCTCAAGTCTTATGTAAGCAAAATATCTTGA	40
QY	1835	tgtgatgataaaccaacttatt	1857
Db	39	TGTGATGATAAACCAAACTTATT	17


```
Best Local Similarity 96.6%; Pred. No. 1.4e-96;
Matches 421; Conservative 0; Mismatches 12; Indels 3; Gaps 2;

QY 1477 agagcaatgtcccaatgcacactgtctcttccaagcgtcgaaacagactgtggatt 1536
|||||
Db 474 AGAGCAATGTGCCAATGCCACTTGTCTTCCAAAGCGTGAACAGACTGCGGGATT 415
|||||

QY 1537 tqcagagtggtttctccatgtcttgaccacaggggtgtgtgtcgtccaggctctagatc 1596
|||||
Db 414 TGCAGAGTGTCTTCCCATGTCCTTGACCACAGGCT-NGTTGCTGCCAGGCTCTAGATC 356

QY 1597 acatgcatcagcgtggggcagaggaatagctattgtctcggcatctctccagggttg 1656
|||||
Db 355 ACATGGCATCAGCGTGGGGCAGAGGCATAGCTATTGTCTCGGGCATCCTTCCAGGGTTG 296
|||||

QY 1657 ggtctcacacaataagaaggctctgctcgtgagttatgtgacgtgctcagccccatgga 1716
|||||
Db 295 GGTCTTACACAATAGAGGCTCTTGTCTGTGATTTATGTGACATGCTCAGCCCCATGGA 236

QY 1717 ctaagcaggggtcgtgtat--aaacactcctggaacgcttgcctcctatccaaatgtt 1774
|||||
Db 235 CTAAGCAGGGGTCTGGGTATAAAAAACACTCTGGAACGCCCTTTGGCCCTGATCCAAATGTT 176
|||||

QY 1775 agcactgctagtgaaactctacttatctcaagttctatgctaaaggcaatttatcttga 1834
|||||
Db 175 AGCACTTGTCTAGTGAAGCTCTACTTATCTCAAGTTCTATGCTAAAGCAATTTATCTTGA 116

QY 1835 tctgatataaaccacacttattagaagatatgcataatataatccataattctcttacc 1894
|||||
Db 115 TGTGATGATAAACCAAACTTATTAGCAAGATATGATATATATTCATTAATATCTTTAC 56

QY 1895 tctgtctccatccttt 1910
|||||
Db 55 TCTGTCTCCATCACTT 40

RESULT 13
AW401870 410 bp mRNA EST 16-FEB-2000
LOCUS
DEFINITION
IMAGE:3053781 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
EST.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 410)
AUTHORS
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
On Jun 22, 1998 this sequence version replaced gi:3246884.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue procurement: Louis M. Staudt, M.D., Ph.D.
DNA Library Preparation: M.B. Soares Lab
DNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.
Location/Qualifiers
1. .410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3053781"
/clone_lib="NIH-MGC_36"
/tissue_type="lymph"
/cell_type="germinal center B cells"
```

```
/cell_line="MGC85"
/lab_host="DH10B (LTI)"
/notes="Vector: p77T3-Pac; Site.1: NotI; Site.2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(0.5-1.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldo, Ph.D., and M. Bento Soares, Ph.D."
BASE COUNT 72 a 154 c 94 g 89 t 1 others
ORIGIN

Query Match 20.0%; Score 382.4; DB 79; Length 410;
Best Local Similarity 99.2%; Pred. No. 2.7e-94;
Matches 394; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1041 tggagcggacgtgcagcagcaggggagggcccccgttcccgcccgagcggcggttgcc 1100
|||||
Db 15 TGGAGCGGAGCGTGCAGGCACAGGGAGGGCCCCGTTCCCGCGCCGGAGCGCGTTGCC 74
|||||

QY 1101 cccgcgccgctgcaggtgtctgaattccctggctccatgcccctctctgaagaccag 1160
|||||
Db 75 CCCC GCCCGCTGCGAGGTGCTGAATCTCCCTGGCTCCATGCCCATCTCTGAAGACCAG 134
|||||

QY 1161 ctgtgaatacgtgagcctctaccaccagcctgcccgcacatgagagacagtgattcaga 1220
|||||
Db 135 CTGTGAATACGTGAGCCTCTACCACCAGCCTGCCGCCCATGATGGAGGACAGTGATTGAGA 194
|||||

QY 1221 tgactacataatgttctcctgcctgcacactccccagctatcccccaacccagcgtcgg 1280
|||||
Db 195 TGACTACATCAATGTTCTTCCCTGCCTGACAACTCCCGAGCTATCCCCCAACCCAGGCTCGGA 254
|||||

QY 1281 ctgtgtgccaaggagctcatctatctgctgagtgatcacaatacctgtcttcattgtctc 1340
|||||
Db 255 CTGTGTGGCAAGGAGTGTCTATCTATCTTCTGATGTCCAAATGCTCTTCATGTGTTCTC 314
|||||

QY 1341 agagcctcatcacttcccctgcccctctgcaactccccagctatcccccaacccagcgtcgg 1400
|||||
Db 315 AGAGCCTCATCACTTCCCATGCCCATCTGAGCTCCCATCCCATCTATCTGT-GCCT 373
|||||

QY 1401 gagcatggtctgtgcccacaggctcgtcttcacacctt 1437
|||||
Db 374 GAGCATGGCTCTTCCCCCAGGTGCTGTTCACACACTT 410
|||||

RESULT 14
AW456707 404 bp mRNA EST 06-JUN-1997
LOCUS
DEFINITION
aai3904.r1 Soares_NhMPu_S1 Homo sapiens cDNA clone IMAGE:813174
5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
EST.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 404)
AUTHORS
Hallier,L., Allen,M., Bowles,L., Dubuque,T., Gelsel,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-Merck EST Project 1997
Unpublished (1997)
COMMENT
On Jan 14, 1998 this sequence version replaced gi:1877745.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amersham
```


Align seg 1/1 to: X28179 from: 1 to: 1047

```

1 MetAspArgTTPLeuTTPProLeuTyrPheLeuProValSerGlyAlaLe 17
|||||
150 ATGGACTTCTGCTTTGGCCACTTTACTTCTGCCAGTATCAGGGGGCCCT 199
17 uArgLLeuProGluValValGluGlyGlySerValT 34
|||||
200 GAGGATCTCCAGAGTAAGGTAGAGGGGAGCTGGGGGATCAGTTA 249
34 hrLeuLeuCysProLeuProGluMetHisValArgLeuTyrLeuCysArg 50
|||||
250 CCATCAAGTGGCCACTTCTGGAATGCAATGATGAGGATATATCTGTGCGG 299
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67
|||||
300 GAGATGGCTGATCGTGAACATGTGTACCGTGTATCCACCCACCACTT 349
67 eileLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgL 84
|||||
350 CATCAAGGCAAGATACAGGGCCGAGTTACTCTGAAGCAATACCCAGCA 399
84 ysAsnLeuPheLeuValGluValThrGlnLeuThrGlnSerAspSerGly 100
|||||
400 AGAATCTGTTCTAGTGGAGGTAAACACAGCTGACAGAAAGTACAGCGGA 449
101 ValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117
|||||
450 GTCTATGCTTGGCGAGC.GGCATGAACACAGACCGGGGAAAGACCCAGAA 498
117 sValThrLeuAsnValHisSerGluTyrGluProSerTrpGluGluGlnP 134
|||||
499 AGTCACCTCAATGTCCACAGTGAATACGAGCCATCATGGAAGAGGAGC 548
134 toMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMet 150
|||||
549 CAATGCTGAGATCCAAATGGTTTCATCTGCCCTATTGTTCCAGATG 598
151 ProAlaTyrAlaSerSerLysPheValThr..ArgValThrThrPro 166
|||||
599 CTTGATATCCCGGTTCTTCCACATTCGTAAACCGCAGAGTTACCAACCA 648
167 Ala.GlnArgGlyLysValProValHisHisSerSerProThrThrG 183
|||||
49 GCTTAAAGGGGCAAGGTCCTCCAGTTCACCACTCTCCGCCACCAACC 698
183 Inile.ThrHisArgPro.ArgValSerArgAlaSerSerValAlaGly 199
|||||
699 AAATTCACCCACCGCCCTTCGAGTNCAGAGCATCTTCAGTAGCAGGTG 748
199 spLysProArg..ThrPheLeuProSerThrThrAlaSerLysLysSerAl 215
|||||
749 ACAAGCCCCGAAACTTCTGCCATCCACTACAGCCCTCAAAATCTCAGC 798
215 aLeuGluGlyLeuLeuLys...ProGlnThrProSerTyrAsnHisHist 231
|||||
799 TCTGGAAGGCTGCTTCAAGCCCCAGAGCCCGCCAGCTACACAN..ACA 847
231 hrArgLeuHisArgGlnArgAlaLeuAsp...Tyr.GlySerGlnSerG 246
|||||
848 CCAGCTGACAGGAGAGAGCACTGGATACTTATGGGNTCACAGTCTGG 897
246 yArgGluGly.....GlnGlyPhe...HisIleLeuIlePro.ThrIle 259
|||||
898 G...GAGGGAANGNCCAAAGGATTTNACATCTCTGATTCGCCGACCATC 944
260 LeuGlyLeuPhe 263
|||||
945 NTGGGGCCCTTT 956

```

seq_name: N_Geneseq_36:T31290

```

seq_documentation_block:
ID T31290 standard; cDNA; 3095 BP.
AC T31290;
DE Mouse poly-immunoglobulin receptor, cDNA.
KW Mouse; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
OS poly; sorbinus; murine; ss.
FH Key Location/Qualifiers
FT cds 85..2400
FT CDS /*tag= a
W09621012-Al.
PN 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hatt AC, Lehner T, Ma JKC;
DR WPI: 96-333987/33.
DR P-PSDB: W03180.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure: Pages 117-121: 152pp; English.
CC The present sequence encodes the mouse poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-608, or esp. residues 13-45, 1-120, 110-230, 320-450,
CC 440-550, 550-606 or 550-627 comprises a pp as above in association with
CC the Ig of the invention comprises a pp as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The pp protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 3095 BP; 861 A; 796 C; 784 G; 654 T;

```

alignment_scores:
Quality: 198.00 Length: 168
Ratio: 1.980 Gaps: 5
Percent similarity: 59.524 Percent identity: 31.548

alignment_block:

US-09-135-238B-2 x T31290 ..

Align seg 1/1 to: T31290 from: 1 to: 3095

```

10 PheLeuProValSerGlyAlaLeuArgLLeuLeuProGluValLysValG 26
|||||
121 TTTTACGGGGTCTCCACAAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
26 uGlyGluLeuGlyGlySerValThrLysCysProLeuProGlu.... 41
|||||
171 TAGTATAGAGGGCGACTCTGTTCCATCATCGTGTACTACCCAGACCT 220
42 .....MetHisValArgLLeuTyrLeuCysArgGluMetAlaGlySer 55
|||||
221 CTGTCACCGGCACACCCGGAATACTGTGTCGCCAAGGAGCC...AGC 267
268 GGCATGTGCACACGCTCATCTCTCAATGGCTACCTCTCCAGGAGTA 317
56 GlyThrCysGlyThrValValSerThrThrAsnPheIleLysAlaGlu 72
|||||
268 GGCATGTGCACACGCTCATCTCTCAATGGCTACCTCTCCAGGAGTA 317
72 rLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeu 89
|||||

```


334 roValProGlyProGlyAlaProLeuProAlaProLeuGlnValSer 350
|||||
1146 CGGTTCCCGCGCGGAGCGCGTTGGCCCGCGCGCTGCAGGTGCT 1195
351 GluSerProTrrpLeuHisAlaProSerLeuLysThrSerCysGluTyrVa 367
|||||
1196 GAATCTCCCTGGCTCCATGCCCATCTCTGAAGACAGCTGGAATACGT 1245
367 lSerLeuTyrHisGlnProAlaAlaMetMetGluAspSerAspSerAspa 384
|||||
1246 GAGCTCTACACACAGCGCTGCCCATGATGAGGACAGTATTGATGATG 1295
384 spTyrIleAsnValProAla 390
|||||
1296 ACTACATCAATGTTCTCTGCC 1315

seq_name: N_Geneseq_36:X28179

seq_documentation_block:

ID X28179 standard; CDNA; 1047 BP.

AC X28179;

DE Human PIGRL-1 coding sequence fragment.

KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;

KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;

KW diagnosis; therapy; ss.

OS Homo sapiens.

PN EP-905238-A2.

PD 31-MAR-1999.

PF 14-AUG-1998; 306487.

PR 30-OCT-1997; US-961564.

PR 25-AUG-1997; US-056935.

PA (SMIK) SMITHKLINE BEECHAM CORP.

PI Sweet RW, Truneh A, Wu S;

DR WPI: 99-192666/17.

DR P-PSDB; Y05002.

PT New polypeptides encoding PIGRL-1 useful for treating diseases such

PT as X-linked Severe Combined Immunodeficiency

PS Disclosure: Page 8; 26pp; English.

CC This sequence encodes the human PIGRL-1 protein of the invention.

CC Autoimmune diseases involving altered expression or activity of PIGRL-1

CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined

CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be

CC diagnosed or susceptibility to them predicted by: (1) determining whether

CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or

CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.

CC Patients deficient in PIGRL-1 can be treated by administering either the

CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.

CC Patients with excessive expression or activity of PIGRL-1 can be treated

CC by administering an antagonist of PIGRL-1, an antisense nucleic acid

CC molecule which inhibits the expression of PIGRL-1 or administering

CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can

CC be used to identify its agonists by contacting a cell expressing PIGRL-1

CC with a candidate compound in the presence of a signal system and noting

CC the candidate as an agonist if a signal is produced. The same method can

CC be used to identify antagonists of PIGRL-1 but the presence of an

CC antagonist is indicated by a decrease in production of the signal.

CC Antibodies against PIGRL-1 may be used to isolate or identify clones

CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to

CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.

CC This information may then be correlated with the incidence of autoimmune

CC disease in those patients to identify whether the mutation causes the

CC disease.

CC

SQ Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T;

alignment_scores:

Quality: 1125.00 Length: 275

Ratio: 4.429 Gaps: 12

Percent Similarity: 92.364 Percent Identity: 90.182

alignment_block:

US-09-135-238B-2 x X28179

34 hrIleYscysProLeuProGluMetHisValArgIleTyrLeuCysArg 50
|||||
246 CCATCAAGTGGCCCATCTCTGAAATGCGATGTGAGGATATATCTGTGCGCGG 295
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67
|||||
296 GAGATGGCTGGATCTGGAACATGTGTGACCTGATCCACCAACCACTT 345
67 eIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgL 84
|||||
346 CATCAAGGCGAATACAAGGGCGGAGTTACTCTGAAGCAATACCCACGCA 395
84 ysAsnLeuPheLeuValLysValThrGlnLeuThrGluSerAspSerGly 100
|||||
396 AGAATCTGTTCTAGTGAGGTAACACAGCTGACAGAAAGTGACAGCGGA 445
101 ValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117
|||||
446 GTCTATGCTGGGAGGGGCGATGAACACAGACAGCGGGGAAAGACCGAA 495
117 sValThrLeuAsnValHisSerGluTyrGluProSerTrrpGluGluGlnP 134
AGTCAACCTGAATGTCCACAGTCAATACGAGCCATCATGGGAAGAGCAGC 545
134 roMetProGluThrProLysTrrpPheHisLeuProTyrLeuPheGlnMet 150
|||||
546 CAATGCTGAGACTCCAAATGTTTTCATCTGCCCTATTGTTCCAGATG 595
151 ProAlaTyrAlaSerSerSerLysPheValThrArgValThrThrProAl 167
596 CTTGCATATGCCAGTTCTTCCAAATTCGTACACAGAGTTTACACACACAGC 645
167 aGlnArgGlyLysValProValHisHisSerSerProThrThrGlnI 184
|||||
646 TCAAGGGGCAAGGTCCTCAGTTCAACCTTCAACCTTCCCTCCACCCCAA 695
184 leThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLys 200
696 TCACCCAGCCGCTCGAGTGTCCAGAGCATCTTCAGTAGCAGGTGACAAG 745
201 ProArgThrPheLeuProSerThrThrAlaSerLysLysSerAlaLeuGl 217
746 CCCCAGAACCTTCTGCCATCCACTACAGCTTCAAAATCTCAGCTCTGGA 795
217 uGlyLeuLeuLysProGlnThrProSerTyrAsnHisHisThrArgLeuH 234
796 GGGGCTGCTCAAGCCAGAGCCGCGCCAGCTACAAACACACACACAGCTGC 845
234 isArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGln 250
845 ACAGGCAGAGACACTGGACTATGGCTCACAGTCTGGGAGGGAGGCCAA 895
251 GlyPheHisIleLeuIleProThrIleLeuGlyLeuPheLeuAlaLe 267
896 GGATTTTCACATCTGATCCCGACCATCTGGGCGCTTTTCTGCTGGCACT 945
267 uLeuGlyLeuValValLysArgAlaValGluArgLysAlaLeuSerA 284
946 TCTGGGGCTGGTGGTGAAGGGCGCTTGAAGGAGGAAAGCCCTCTCCA 995
284 rgArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln 300
996 GGGGGCCCGCGAGCTGGCGGTGAGGATCGCGCCCTGGAGAGCTCCAC 1045
301 ArgProArgGlySerProArgProArgSerGlnAsnAsnIleTyrSerAl 317
1046 AGGCCCGCGGGTGCCTGCGGAGCGGCTCCCAAAACACATCTACAGCCG 1095
317 aCysProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlaP 334
1096 CTGCGCCGCGCGCTCTGTGGAGCGGACGCTGACGACAGGAGGCGCC 1145



OM of: US-09-135-238B-2 to: N_Geneseq_36:* out_format : pfs

Date: Sep 12, 2000 6:55 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frank_pdn.model -DEV=xlip  
-Q=/cygn2_1/USPTO_SPOOL/US09135233/runat_24082000_091324_17918/app_query.fasta_1.1282  
-DB=N_Geneseq_36 -FMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM-ext -MNTLEN=0 -MAXLEN=100000  
-USER=US09135238.ecgn1_123 -NCPV=6 -ICPU=3 -LONGLOG -NO_XLPYX  
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-135-238B-2  
Query length: 390  
Database: N_Geneseq_36:*  
Database sequences: 311585  
Database length: 125096042  
Search time (sec): 121.040000
```

score_list:

Sequence	Strd Orig	zScore	EScore	Len	Documentation
N_Geneseq_36:X28178	2047.00	1129.34	5.3e-111	2040	Human PI6RL-1 coding sequence.
N_Geneseq_36:X28179	1125.00	1174.09	8.5e-58	1047	Human PI6RL-1 coding sequence.
N_Geneseq_36:X12190	198.00	199.21	0.0017	3095	Mouse poly-immunoglobulin rece
N_Geneseq_36:X15422	188.00	170.27	0.0657	29392	Mouse poly Ig receptor protei
N_Geneseq_36:X20383	187.50	192.24	0.0042	1911	cDNA for human immunity relate
N_Geneseq_36:X31291	185.50	185.74	0.0096	3269	Rat poly-immunoglobulin recept
N_Geneseq_36:X30857	185.00	189.53	0.0059	1936	Secretory component DNA fragme
N_Geneseq_36:X07407	182.00	186.82	0.0083	1839	Human secretory immunoglobulin
N_Geneseq_36:X30856	182.00	186.01	0.0093	2031	Secretory component DNA fragme
N_Geneseq_36:X31288	182.00	183.02	0.0136	2919	Human poly-immunoglobulin rece
N_Geneseq_36:X31289	182.00	181.23	0.0171	3630	Bovine poly-immunoglobulin rece
N_Geneseq_36:X96033	171.00	174.33	0.0414	2084	Rat kidney injury related mole
N_Geneseq_36:X96032	171.00	172.62	0.0515	2566	Rat kidney injury related mole
N_Geneseq_36:X14498	166.50	170.51	0.0676	1876	Clone 1 for truncated poly Ig-
N_Geneseq_36:X31287	166.50	165.34	0.1312	3517	Rabbit poly-immunoglobulin rec
N_Geneseq_36:X28250	159.50	167.34	0.1015	1137	Human PI6RL-2 coding sequence.
N_Geneseq_36:X14499	145.00	148.11	1.20	1876	Clone 2 for truncated poly Ig-
N_Geneseq_36:X18694	143.50	133.93	7.37	8690	Plasmid pD17-huM4.H1. Inhibit
N_Geneseq_36:X24074	142.00	150.90	0.8361	914	Human EST R2810 DNA. Complexes
N_Geneseq_36:X18598	139.50	130.12	12.02	8321	Plasmid expressing hBR96-A. In
N_Geneseq_36:X44439	138.50	146.63	1.44	985	Mycobacterium tuberculosis anti
N_Geneseq_36:X64548	138.50	146.63	1.44	985	M. tuberculosis immunogenic pol
N_Geneseq_36:X87689	136.50	149.12	1.05	565	EST clone EN260. New polynucle
N_Geneseq_36:X90988	136.50	140.14	3.32	1683	Human dmpase gene 5' region.
N_Geneseq_36:X43844	136.00	123.79	27.06	11529	Plasmid pRH4602. Antibody con
N_Geneseq_36:X43846	135.00	123.35	28.60	10704	Plasmid pRH4625. Antibody con
N_Geneseq_36:X43847	131.00	118.16	55.71	12132	Plasmid pRH4607. Antibody con
N_Geneseq_36:X35099	130.00	133.70	7.59	1617	Antibody B heavy chain. Prodn.
N_Geneseq_36:X62510	129.50	134.18	7.13	1431	Primaetised anti-human B7.1 an
N_Geneseq_36:X35485	129.50	134.18	7.13	1431	Macaque primaetized 7C10 heavy
N_Geneseq_36:X06877	127.00	128.05	15.66	2197	ISI096 transposon. Leucine aux
N_Geneseq_36:X39291	126.50	120.88	39.26	4926	Plasmid pHC65 nucleotide sequ
N_Geneseq_36:X43497	126.50	114.03	94.63	11336	Approximate nucleotide sequ
N_Geneseq_36:X49944	125.50	129.22	13.48	1576	Human anti-HBs heavy chain. Hu
N_Geneseq_36:X62937	125.50	122.89	30.35	3400	3F4 human G2/G4 chimeric anti
N_Geneseq_36:X29059	125.50	116.64	67.64	7266	Human calcium channel a1b subu
N_Geneseq_36:X15932	124.50	116.45	69.37	6357	Anti-IgE VH expression vector.
N_Geneseq_36:X62175	124.50	108.70	187.39	16812	HSV-2 strain SB5 Contig ID 12
N_Geneseq_36:X11459	124.00	125.90	20.63	1951	C. acidovorans gamma-lactamase
N_Geneseq_36:X12632	124.00	124.32	25.27	2364	CD4-specific CDR-grafted heavy

N_Geneseq_36:T50962 + 124.00 115.30 80.34 7073 ! TF8-5G9 CDR-grafted heavy c
N_Geneseq_36:T57472 + 123.50 125.47 21.80 1929 ! Sorghum bicolor (L.) Moench
N_Geneseq_36:T96035 + 123.00 125.54 21.60 1795 ! Human kidney injury related
N_Geneseq_36:T72685 + 123.00 113.19 105.27 8051 ! Sugar biosynthesis gene clu
N_Geneseq_36:X25774 + 123.00 113.08 106.78 8160 ! S.erythraea erythromycin-sy

seq_name: N_Geneseq_36:X28178

seq_documentation_block:

ID X28178 standard; CDNA; 2040 BP.
AC X28178;
DT 16-JUN-1999 (first entry)
DE Human PI6RL-1 coding sequence.
KW PI6RL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;
KW X-linked Severe Combined Immunodeficiency; XSCID; IGA deficiency;
KW diagnosis; therapy; ss.
OS Homo sapiens.
PN EP-905238-A2.
PP 31-MAR-1999. 306487.
FD 14-AUG-1998; US-961564.
PR 30-OCT-1997; US-056935.
PR 25-AUG-1997; US-056935.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Sweet RW, Truneh A, Wu S;
WP1: 99-192666/17.
P-PSDB: Y05001.
DR New polypeptides encoding PI6RL-1 useful for treating diseases such
PT as X-linked Severe Combined Immunodeficiency
PT Claim 4; Page 18-19; 26pp; English.
PS This sequence encodes the human PI6RL-1 protein of the invention.
CC Autoimmune diseases involving altered expression or activity of PI6RL-1
CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined
CC Immunodeficiency (XSCID) and IGA deficiency. These diseases can be
CC diagnosed or susceptibility to them predicted by: (1) determining whether
CC there is a mutation in the genomic copy of the gene encoding PI6RL-1; or
CC (2) measuring the amount of PI6RL-1 in a sample derived from the patient.
CC Patients deficient in PI6RL-1 can be treated by administering either the
CC PI6RL-1 DNA or its complement or an agonist of PI6RL-1 to the patient.
CC Patients with excessive expression or activity of PI6RL-1 can be treated
CC by administering an antagonist of PI6RL-1, an antisenesc nucleic acid
CC molecule which inhibits the expression of PI6RL-1 or administering
CC sufficient PI6RL-1 to compete with the endogenous activity. PI6RL-1 can
CC be used to identify its agonists by contacting a cell expressing PI6RL-1
CC with a candidate compound in the presence of a signal system and noting
CC the candidate as an agonist if a signal is produced. The same method can
CC be used to identify antagonists of PI6RL-1 but the presence of an
CC antagonist is indicated by a decrease in production of the signal.
CC Antibodies against PI6RL-1 may be used to isolate or identify clones
CC expressing PI6RL-1. Polynucleotides encoding PI6RL-1 may be used to
CC identify chromosomal mutations in the gene encoding PI6RL-1 in patients.
CC This information may then be correlated with the incidence of autoimmune
CC disease in those patients to identify whether the mutation causes the
CC disease.

Sequence 2040 BP; 478 A; 616 C; 494 G; 452 T;

alignment_scores:

Quality: 2047.00 Length: 390
Ratio: 5.252 Gaps: 0
Percent Similarity: 99.744 Percent Identity: 99.744

alignment_block:

US-09-135-238B-2 x X28178

Align seg 1/1 to: X28178 from: 1 to: 2040

1 MetAspArgTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLe 17
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
146 ATGCAGCTCTGGCTTTGGCCACTTACTTCTCCGAGTATCAGGGGCCCT 195
17 uArgIleLeuProGluValIysValGluGluLeuGlyGlySerValt 34
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
196 GAGGATCTCTCCAGAGTAAGTAGAGGGGGGAGCTGGCGGATCAGTTA 245



```

* 31620 32323: contig of 704 bp in length
* 32324 32423: gap of 100 bp
* 32424 33110: contig of 687 bp in length
* 33111 33210: gap of 100 bp
* 33211 33902: contig of 692 bp in length
* 33903 34002: gap of 100 bp
* 34003 34708: contig of 706 bp in length
* 34709 34808: gap of 100 bp
* 34809 35500: contig of 692 bp in length
* 35501 35600: gap of 100 bp
* 35601 36300: contig of 700 bp in length
* 36301 36400: gap of 100 bp
* 36401 37097: contig of 697 bp in length
* 37098 37197: gap of 100 bp
* 37198 37900: contig of 703 bp in length
* 37901 38000: gap of 100 bp
* 38001 38695: contig of 695 bp in length
* 38696 38795: gap of 100 bp
* 38796 39487: contig of 692 bp in length
* 39488 39587: gap of 100 bp
* 39588 40291: contig of 704 bp in length
* 40292 40391: gap of 100 bp
* 40392 41101: contig of 710 bp in length
* 41102 41201: gap of 100 bp
* 41202 41886: contig of 685 bp in length
* 41887 41986: gap of 100 bp
* 41987 42686: contig of 700 bp in length
* 42687 42786: gap of 100 bp
* 42787 43484: contig of 698 bp in length
* 43485 43584: gap of 100 bp
* 43585 44276: contig of 692 bp in length
* 44277 44376: gap of 100 bp
* 44377 45078: contig of 702 bp in length
* 45079 45178: gap of 100 bp
* 45179 45869: contig of 691 bp in length
* 45870 45969: gap of 100 bp
* 45970 46664: contig of 695 bp in length
* 46665 46764: gap of 100 bp
* 46765 47456: contig of 692 bp in length
* 47457 47556: gap of 100 bp
* 47557 48249: contig of 693 bp in length
* 48250 48349: gap of 100 bp
* 48350 49042: contig of 693 bp in length
* 49043 49142: gap of 100 bp
* 49143 49843: contig of 701 bp in length
* 49844 49943: gap of 100 bp
* 49944 50640: contig of 697 bp in length
* 50641 50740: gap of 100 bp
* 50741 51438: contig of 698 bp in length
* 51439 51538: gap of 100 bp
* 51539 52231: contig of 693 bp in length
* 52232 52331: gap of 100 bp
* 52332 53020: contig of 689 bp in length
* 53021 53120: gap of 100 bp
* 53121 53857: contig of 737 bp in length
* 53858 53957: gap of 100 bp
* 53958 54652: contig of 695 bp in length
* 54653 54752: gap of 100 bp

```

alignment_scores:

```

Quality: 284.00      Length: 80
Ratio: 4.733        Gaps: 2
Percent Similarity: 75.000    Percent Identity: 73.750

```

alignment_block:

```

US-09-135-238b-2 x AC027719/rev

```

```

Align seg 1/1 to reverse of: AC027719 from: 1 to: 58628

```

```

2 AspArgTrpLeuTrp..... 6
||| ||| |||
40676 GACGCTGGAGCTGGGAAGTAGATGGGCTGGGATAGTAGATAGTAGATAGG 40627

```

```

7 .....ProLeuTyrPheLeuPro.....ValSerGlyAlaLeuArgI 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40626 GTCCTCTGCAATTAATCTCTCCCTGCAGATCAGGGCCCTGAGGA 40577

19 leLeuProGluValValGluGluGluLeuGlySerValThrIle 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40576 TCCTCCCAAGAAAGTAGAGGGGAGCTGGCGGATCAGTTACCATC 40527

36 LysCysProLeuProGluMethHisValArgIleTyrLeuCysArgGluMe 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40526 AAGTGGCCACTTCTCTGAATGATGTGAGGATATATCTGTGCCGGAGAT 40477

52 talaGlySerGlyThrCysGlyThrValValSerThrThr 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40476 GGCTGGATCTGGAACATGTGTACCGTGTATCCACCACC 40437

seq_name: gb_ro:MMU06431

seq_documentation_block:
LOCUS MMU06431 3095 bp mRNA ROD 26-MAY-1995
DEFINITION Mus musculus B6/CBAF1J polymeric immunoglobulin receptor mRNA,
complete cds.
ACCESSION U08431
VERSION 008431.1 GI:458421
KEYWORDS mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Piskurich,J.F., Blanchard,M.H., Youngman,K.R., France,J.A. and
Kaetzel,C.S.
TITLE Molecular cloning of the mouse polymeric Ig receptor. Functional
regions of the molecule are conserved among five mammalian species
J. Immunol. 154 (4), 1735-1747 (1995)
J5139517
2 (sites)
Piskurich,J.F., Hsieh,M., Youngman,K.R. and Kaetzel,C.S.
AUTHORS Molecular cloning of mouse polymeric Ig receptor
J. Immunol. 150, 38A-38A (1993)
J3 (bases 1 to 3095)
REFERENCE Piskurich,J.F.
AUTHORS Molecular Cloning and Regulation of the Polymeric Immunoglobulin
Receptor
JOURNAL Thesis (1994) Pathology, Case Western Reserve University
REFERENCE 4 (bases 1 to 3095)
AUTHORS Piskurich,J.F.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-1994) Janet F. Piskurich, Case Western Reserve
University, Pathology, Biomedical Research Building, Cleveland, OH
44060, USA
FEATURES
Location/Qualifiers
1..3095
/organism="Mus musculus"
/strain="B6/CBAF1J (C57 Black 6 x CBA)"
/db_xref="taxon:10090"
/clone="G20"
/clone_lib="catalog #935302, Stratagene Cloning Systems,
La Jolla, CA"
/sex="female"
/tissue_type="liver"
<1..84
85..138
85..2400
/codon_start=1
/function="binds and transports polymeric immunoglobulin"
/product="polymeric immunoglobulin receptor"
/protein_id="AAA67440.1"
/db_xref="GI:458422"
/translation="MRLVLTLLTVFSGVSTKSPFGPOEVSSTEGDSVSTCYVDP
TSVNRHTRKVMCRGASCMCTLLSSNGVLSKYSGRANLFPENNTFVNIQLTQ
DGTGSGCLGTSNGLSFDSPVDELPSPDRHYTMDIGRVTIECPKRENP

```


34 roArgProArgSerGlnAsnAlaMetTyrSerAlaCysProArgAla 50
 1062 CGCAGCCGCGCTCCCAAAACATCTACAGCGCTGCGCGCGCT 1111
 51 ArgGlyAlaAspAlaLaglyThrGlyGluAlaProValProGlyProGly 67
 1112 CGTGGAGCGACGCTGCAGGACAGGGAGGCCCGCTTCCGCGCCCG 1161
 67 yAlaProLeuProProAlaProLeuGlnValSerGluSerProTyrLeuH 84
 1162 AGCGCCGCTGCGCCCGCGCGCTGAGGTCTGAATCTCCCTGGCTCC 1211
 84 iSAlaProSerLeuLysThrSerCysGluTyrValSerLeuTyrHisGln 100
 1212 ATGCCCATCTCTGAAGACAGCTGTGAATACGTGAGCCTCTACACAC 1261
 101 ProAlaAlaMetMetGluAspSerAspAspTyrIleAsnValPr 117
 1262 CTGCGCGCCATGATGGAGACAGTGTATTCAGATGACTACATGATGTC 1311
 117 oala 118
 1312 TGCC 1315

seq_name: N_Geneseq_36:T93610

seq_documentation_block:

ID T93610 standard; DNA; 3946 BP.

AC T93610; 1998 (first entry)

DE Mycobacterium tuberculosis genomic DNA fragment (I).

KW Tuberculosis; mycobacteria; infection; diagnosis;

OS antimycobacterial; antibiotic; vaccine; ss.

PN WO9741252-A2.

PD 06-NOV-1997.

PF 18-APR-1997; E01973.

PR 29-APR-1996; DE-017184.

PA (GFBF) GBF GES BIOTEC FORSCHUNG GMBH.

PI Espitia C. Honisch C. Moreno C. Singh M;

DR WPI; 97-549750/50.

DR P-PSDB: W31852-55.

PT New DNA and related proteins or RNA derived from M. tuberculosis -
 used for diagnosis of mycobacterial infections, monitoring
 PT vaccination and development of anti-mycobacterial agents
 PS Claim 1; Fig 9; 55pp; English.

CC New claimed DNA (A) has one of 3 isolated Mycobacterium
 CC tuberculosis DNA sequences of 3946 bp (I) (T93610), 2653 bp (VI)
 CC (see T93611) or 440 bp (IX) (see T93612), or hybridises to, or is a
 CC fragment of these sequences. Also claimed are RNA transcripts
 CC of (A), and proteins encoded by (A). To isolate (I), a cosmid
 CC library of M. tuberculosis DNA was screened with degenerate probes
 CC (see T93621) containing GC-rich regions and designed to encode part
 CC of a proline-rich protein. (I) contains open reading frames (ORFs)
 CC for 4 proteins (see W31852-55). These ORFs can be amplified and
 CC cloned into vectors for expression in E. coli or other hosts.
 CC Clones (VI) and (IX) were obtained from a phage library and
 CC together encode 3 proteins (see W31851 and W31856-57). The claimed
 CC DNA sequences can be used for diagnosing tuberculosis and other
 CC mycobacterial infections in humans or animals, for identifying
 CC mycobacteria in (clinical) samples by hybridisation or
 CC as for epidemiological studies, for differentiation between strains, as well
 CC the development of anti-mycobacterial drugs and vaccines.
 SQ Sequence 3946 BP; 518 A; 1541 C; 1345 G; 542 T;

alignment_scores:
 Quality: 101.50 Length: 107
 Ratio: 1.750 Gaps: 6
 Percent Similarity: 54.206 Percent Identity: 36.449

alignment_block:

US-09-135-238B-2_COPY_273_390 x T93610 ..

Align seg 1/1 to: T93610 from: 1 to: 3946

1 LysArgAlaValGluArgArgLys.....Al 9
 1617 GCGCGTTCGCGCGCGCGCGCGCGCGCTTCTGGATGCTGTGAAGT 1666
 9 aLeuSerArgAlaArgAlaArgLeuAlaValArgMetArgAlaLeuGluS 26
 1667 GCTCCCGCGCGTGCCTCCCGCGCGCGCGCGCGCGCGCGCGAGCGAG 1716
 26 erSerGlnArgProArgGlySerProArgProArgSerGlnAsnAla 42
 1717 TAGGCGCGCTTCCCGCGCGTTCGCGCGCGCGCGCGCGCGCGAGGAGCTG 1766
 43 TyrSerAlaCysProArgArgAlaArg.GlyAlaAspAla.....AlaG 57
 1767 CGCCACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGGAGAGTAGGCG 1816
 57 lyThrGlyGluAlaProVal.....ProGlyPro..... 66
 1817 GCGTTCG 1866
 67GlyAla.ProLeuProProAlaProLeuGlnValSerGluSerP 81
 1867 GCG 1916
 81 roTrpLeuHisAlaPro 86
 1917 GGTTCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1933

seq_name: N_Geneseq_36:V88273

seq_documentation_block:

ID V88273 standard; CDNA; 266 BP.

AC V88273;

DT 12-FEB-1999 (first entry)

DE EST clone EAL10.

KW Expressed sequence tag; secreted protein; haematopoiesis regulator;

KW tissue growth; activin; inhibitor; tumour invasion suppressor; EST; human;

KW chemotaxis; chemokines; haemostasis; gene therapy; thrombolysis;

KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

OS Homo sapiens.

PN WO9845437-A2.

PD 15-OCT-1998.

PF 10-APR-1998; U06956.

PR 10-APR-1997; US-837312.

PA (GEMY) GENETICS INST INC.

PI Racie LA, Spaulding V, Treacy M;

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,

DR WPI; 99-070078/05.

CC New polynucleotides encoding human secreted proteins - derived from

CC e.g. human blood, kidney, foetal lung, placenta, testes, brain,

CC ovary, pituitary, retina and colon cDNA libraries

PS Claim 1; Page 341; 64pp; English.

CC The present sequence represents an expressed sequence tag (EST), and is
 CC a polynucleotide of the invention. The polynucleotides of the invention
 CC are all secreted EST sequences and proteins encoded by them are predicted to
 CC sources. The EST sequences and proteins encoded by them are suitable for
 CC having useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity, haemostatic
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The EST sequences are also stated to be useful for gene
 CC therapy.

SQ Sequence 266 BP; 50 A; 56 C; 114 G; 46 T;

OM of: US-09-135-238B-2_COPY_273_390 to: N_Geneseq_36:* out_format : pfs

Date: Sep 12, 2000 6:56 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O/cgn2.1/USPTO.spool/US09135238/runat_24082000_091324_17918/app_query.fasta_1.1282
-Da-N_Geneseq_36 -OPM=fastap -SUFFIX=p2n.rng -GAPOP=12,000
-GAPEXT=4,000 -MINMATCH=0,100 -LOOPEXT=0,000 -LOOPEXT=0,000
-GAPOP=4,500 -GAPEXT=0,050 -XGAPOP=10,000 -XGAPEXT=0,500
-XGAPOP=6,000 -XGAPEXT=7,000 -XGAPOP=10,000 -XGAPEXT=0,500
-DEL0P=6,000 -DEL0P=7,000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=1000000
-USER=US09135238_cgn1_1_123 -NCPV=6 -ICPV=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-135-238B-2_COPY_273_390

Query length: 118

Database: N_Geneseq_36:*

Database sequences: 311585

Database length: 125096042

Search time (sec): 121.040000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
N_Geneseq_36:X28178	+	616.00	734.99	2040	Human PIGRL-1 coding sequence.
N_Geneseq_36:T93610	+	101.50	114.54	88.60	Mycobacterium tuberculosis gen
N_Geneseq_36:T93610	+	100.00	133.41	7.87	EST clone EAL10. New polynucle
N_Geneseq_36:T93610	+	100.00	118.33	54.45	Encodes human Retinoic Acid Re
N_Geneseq_36:T93610	+	100.00	118.33	54.45	Human retinoic acid receptor g
N_Geneseq_36:T93610	+	96.50	115.09	82.56	Human retinoic acid receptor g
N_Geneseq_36:T93610	+	96.00	117.41	61.31	31-O-desmethyl-FK506 O-methyl
N_Geneseq_36:T93610	+	96.00	113.27	104.27	Nucleotide sequence of human
N_Geneseq_36:T93610	+	95.50	113.62	99.62	hNET cDNA sequence from p1 c
N_Geneseq_36:T93610	+	95.00	113.62	99.62	cDNA encoding human netrin. Hu
N_Geneseq_36:T93610	+	95.00	92.81	1.4e+03	Streptomyces frenolicin gene
N_Geneseq_36:T93610	+	95.00	92.81	1.4e+03	Streptomyces roseofulvus fren
N_Geneseq_36:T93610	+	94.50	113.37	102.96	Human Sox1 encoding cDNA. New cl
N_Geneseq_36:T93610	+	94.00	119.54	46.62	hRAR-gamma A clone cDNA. New cl
N_Geneseq_36:T93610	+	94.00	87.02	3.0e+03	Platenolide synthase gene clu
N_Geneseq_36:T93610	+	94.00	87.02	3.0e+03	Platenolide synthase gene clu
N_Geneseq_36:T93610	+	93.50	115.57	77.65	Mouse CAAAX processing enzyme R
N_Geneseq_36:T93610	+	93.50	114.51	88.89	Mouse CAAAX processing enzyme R
N_Geneseq_36:T93610	+	93.00	107.20	227.06	S. avermitilis bkd gene region
N_Geneseq_36:T93610	+	92.00	122.15	33.39	Human B2S2 repressor factor (EP
N_Geneseq_36:T93610	+	91.50	110.41	150.54	Human CAAAX processing enzyme R
N_Geneseq_36:T93610	+	91.50	109.74	163.86	Human CAAAX processing enzyme R
N_Geneseq_36:T93610	+	91.50	109.72	164.28	Human RCBI (hRCBI) polypeptide
N_Geneseq_36:T93610	+	91.50	89.88	2.1e+03	Polynucleotide sequence from
N_Geneseq_36:T93610	+	91.00	110.18	185.01	Streptomyces lividans XP55 gen
N_Geneseq_36:T93610	+	91.00	108.76	185.90	Alcaligenes bronchisepticus an
N_Geneseq_36:T93610	+	91.00	104.30	329.59	Genomic DNA encoding human CNE
N_Geneseq_36:T93610	+	91.00	101.40	478.02	DNA of hspEco-1 encoding the
N_Geneseq_36:T93610	+	91.00	101.40	478.02	Human C-type natriuretic pepti
N_Geneseq_36:T93610	+	90.50	106.59	245.45	Sequence encoding high density
N_Geneseq_36:T93610	+	90.50	105.36	286.67	HSV-2 strain 8S5 Contig ID 11
N_Geneseq_36:T93610	+	90.50	104.31	329.11	HSV-2 strain 8S5 Contig ID 90
N_Geneseq_36:T93610	+	90.00	107.13	229.27	PRV 28K gene. Mutant pseudo-ra
N_Geneseq_36:T93610	+	90.00	86.83	3.1e+03	Streptomyces frenolicin gene
N_Geneseq_36:T93610	+	90.00	86.83	3.1e+03	Streptomyces roseofulvus fren
N_Geneseq_36:T93610	+	89.50	102.09	437.09	Receptor type tyrosine kinase
N_Geneseq_36:T93610	+	89.50	100.63	527.17	Non-B, non-C, non-G hepatitis
N_Geneseq_36:T93610	+	89.50	99.95	575.16	Human adeno-associated virus
N_Geneseq_36:T93610	+	89.00	117.19	63.08	DNA fragment containing the am
N_Geneseq_36:T93610	+	89.00	114.79	85.81	DNA fragment of PAPO-G encodes

N_Geneseq_36:003224 + 89.00 111.55 129.91 828 ! DNA fragment of PAPO-SX enco
N_Geneseq_36:003223 + 89.00 111.39 132.69 846 ! DNA fragment of PAPO-SX enco
N_Geneseq_36:003220 + 89.00 110.95 140.39 896 ! DNA fragment of PAPO-GMCSF e
N_Geneseq_36:003221 + 89.00 110.87 141.78 905 ! DNA fragment of PAPO-GMCSF e
N_Geneseq_36:V11747 - 89.00 107.30 224.27 1443 ! Human MNTF-1-1443 DNA fragm

seq_name: N_Geneseq_36:X28178

seq documentation block:

ID X28178 standard; cDNA; 2040 BP.
AC X28178;
AT 16-JUN-1999 (first entry)
DE Human PIGRL-1 coding sequence.
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;
KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;
KW diagnosis; therapy; ss.
OS Homo sapiens.
PN EP-905238-A2.
PD 31-MAR-1999.
PF 14-AUG-1998; 306487.
PR 30-OCT-1997; US-961564.
PR 25-AUG-1997; US-056935.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Sweet RW, Truneh A, Wu S;
DR WPI: 99-192666/17.
DR P-PSDB: Y05001.
DR New polypeptides encoding PIGRL-1 useful for treating diseases such
PT as X-linked Severe Combined Immunodeficiency
PS Claim 4; Page 18-19; 26pp; English
CC This sequence encodes the human PIGRL-1 protein of the invention.
CC Autoimmune diseases involving altered expression or activity of PIGRL-1
CC may include hyper-IgM immunodeficiency (HIM), X-linked Severe Combined
CC immunodeficiency (XSCID) and IgA deficiency. These diseases can be
CC diagnosed or susceptibility to them predicted by: (1) determining whether
CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or
CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.
CC Patients deficient in PIGRL-1 can be treated by administering either the
CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.
CC Patients with excessive expression or activity of PIGRL-1 can be treated
CC by administering an antagonist of PIGRL-1, an antisense nucleic acid
CC molecule which inhibits the expression of PIGRL-1 or administering
CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can
CC be used to identify its agonists by contacting a cell expressing PIGRL-1
CC with a candidate compound in the presence of a signal system and noting
CC the candidate as an agonist if a signal is produced. The same method can
CC be used to identify antagonists of PIGRL-1 but the presence of an
CC antagonist is indicated by a decrease in production of the signal.
CC Antibodies against PIGRL-1 may be used to isolate or identify clones
CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to
CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.
CC This information may then be correlated with the incidence of autoimmune
CC disease in those patients to identify whether the mutation causes the
CC disease.
SQ Sequence 204C BP; 478 A; 616 C; 494 G; 452 T;

alignment_scores:

Quality: 616.00 Length: 118
Ratio: 5.220 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-135-238B-2_COPY_273_390 x X28178 ..
Align seg 1/1 to: X28178 from: 1 to: 2040

1 LysArgAlaValGluArgGlyAlaLeuSerArgArgAlaArgLe 17
962 AAAAGGGCCGTGAAGGAGGAAGGCTCTCCAGCGGCGCCCGCAGT 1011
17 uAlaValArgMetArgAlaLeuGluSerGlnArgProArgGlySerP 34
1012 GCGCGTGAGGATGCGCGCCCTGGAGACTCCAGAGCCCGCGGGTGC 1061

Db 423 GCTAAAGG--AATTACTTGATGTGATGAATAACAACCTTATTANAGATATNATATATA 479

Search completed: September 12, 2000, 13:57:06
Job time: 13667 sec

High quality sequence stop: 390.

FEATURES

Location/Qualifiers
 1. .404
 /organism="Homo sapiens"
 /db_xref="GDB:6044084"
 /db_xref="taxon:9606"
 /clone="IMAGE:813174"
 /tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
 /lab_host="DH10B"
 /note="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2N8HM, pregnant uterus N8HP0, and fetal heart N8HH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."
 71 a 153 c 114 g 66 t

BASE COUNT

ORIGIN

Query Match 19.9%; Score 380; DB 26; Length 404;

Best Local Similarity 99.5%; Pred. No. 1.2e-93;

Matches 402; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 912 aagccctctccaggcgccgcgactggcgtgaggatggcgccctggagagctccc 971

Db 2 AAGCCCTCTCCAGCGCGCGCGGCGGCTGGCGGTGAGGATGGCGCCCTGGAGAGCTCCC 61

QY 972 agagcccgccggtcgccgacgcgcctcccaaaacatctacagcgctcccgcc 1031

Db 62 AGAGCGCCCGCGGGTGCCTCG- GACCGCGCTCCCAACACATCTACAGCGCTGCCCGC 120

QY 1032 ggcgcgtc-tggagcgagcgtgcaggcacaggaggcccgctcccgcccgag 1090

Db 121 GGGCGCTGCTGGAGCGGACGCTGCGAGGACAGAGGAGCGCCGCTTCCGCGCGCGAG 180

QY 1091 cgcgctgcccccgcccgctgcaggtgtctgaatctccctggctccatgcccatctc 1150

Db 181 CGCCGTTGCGCCCGCGCGCTGCAGGNGTCTGAATCTCCCTGGCTCCATGCCCATCTC 240

QY 1151 tgaagaccagctgtgaatacgtgagcctctaccacagcctgcgcgcatgatggagaca 1210

Db 241 TGAAGACCAAGCTGTGAATAGTGAGGCTCTACCAACAGCCTGCCGCCATGATGGAGACA 300

QY 1211 gtgattcagatgactacaatgttctcgtcctgcctgcacaaactcccagctatcccccaacc 1270

Db 301 GTGATTTCAGATGACTACATCAATGTCTCTGCTGACAACTCCCGAGCTATCCCGCAACC 360

QY 1271 cagcgtcagactgtgtgccaaggatctctatctgctaat 1314

Db 361 CAGCCTCGACTGTGGTCCAGAGAGTCTCATCTACTCTGAT 404

RESULT 15

W01276

LOCUS

DEFINITION za40f12.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone

IMAGE:295055 5', mRNA sequence. 18-APR-1996

W01276

VERSION W01276.1 GI:1273256

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 528)

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,

Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
 Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
 Treviski,E., Waterston,R., Williamson,A., Wohlmann,P. and
 Wilton,R.
 The WashU-Merck EST Project
 Unpublished (1995)
 On Jan 7, 1998 this sequence version replaced gi:947771.

TITLE
JOURNAL
COMMENT

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: Etrprimer

High quality sequence stop: 451.

FEATURES

source

1. .528

/organism="Homo sapiens"

/db_xref="GDB:123976"

/db_xref="taxon:9606"

/clone="IMAGE:295055"

/clone_lib="Soares fetal liver spleen 1NFLS"

/sex="male"

/dev_stage="20 week-post conception fetus"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)

with a modified polylinker; Site.1: Site.2: Eco RI;

1st strand cDNA was primed with a Pac I - oligo(dT) primer

[5' AAGTGAAGAATTAATAAGATCTTTTTTTTTTTTTTTT 3',

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Pac I and cloned into the Pac I

and Eco RI sites of the modified pT7T3 vector. Library

went through one round of normalization. Library

constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT

ORIGIN

Query Match 19.6%; Score 374; DB 90; Length 528;

Best Local Similarity 91.6%; Pred. No. 5.8e-92;

Matches 439; Conservative 0; Mismatches 33; Indels 7; Gaps 4;

QY 1396 gcctgaagcagctgtgccccagctgtgtttcacacaccttgagcagccccctatgtg 1455

Db 6 GCCCTGAGCATGGCTGTGCCCGCCAGGCTGCTTGCACACCTTGGCAGCCCCCTGTAGTTG 65

QY 1456 acaggtaagctgtaggcattgagcaattgtcccaatgcccactgtctcttcccaagc 1515

Db 66 ACAGGTAAGCTGTAGGCATGTAGAGCAATTGTCCCAATGCGACTTGCTTCTTCCAAAGC 125

QY 1516 cgtcgacagactgtgggattgtcagagtggtttcttccatgcttttgaccacaggggtt 1575

Db 126 CGTCGACAGACTGTGGGATTTGCAGAGTGTCTTCCATGCTTTGACACACAGGT-TG 184

QY 1576 gttgtgcaggctctagatcatgcatcagctggtgggagagagcatagctattgtct 1635

Db 185 TTGCTCCCCAGCATCTAGATCAGATGCGCATGAGCTGGGGCAGAGGCATAGCTATTGTCT 244

QY 1636 cgggcatcttccccaggggtgggtcttaccacaaatagaaggctcttgcctgagttatgt 1695

Db 245 CGGGCATCTTCCCAGGGTGGGTCTTTACACAAATAGAGGGCTTCTGTCTGAGTTATGT 304

QY 1696 gaagtgcctcaagcccatgactaagcaggggtctgtat--aaacactctggaacgc 1753

Db 305 GACATGCTCAGCCCCCATGAGCTAAGCAGGGGTCTGGTATATAAAMACACTCTCGAAACGC 364

QY 1754 ctttgcctgatcccaaatgttagcactgttagtgaacgtctacttctcaagtctat 1813

Db 365 CTTTGGCCCTGATCCAAATGTTAGCACTTGTCTAGTAGTGAACGTCTACTTA--NTCAAGTCTAT 422

QY 1814 gctaaaggcaatttatctctgtatgtgatataaaccacaaactatttagcaatagcata 1872

```

alignment_scores:      Quality: 96.50      Length: 86
                       Ratio: 1.969      Gaps: 5
                       Percent Similarity: 56.977      Percent Identity: 33.721

alignment_block:
US-09-135-238B-2_COPY_273_390 x T90988/rev ..
Align seq 1/1 to reverse of: T90988 from: 1 to: 168

```

alignment_scores:
 Quality: 100.00 Length: 76
 Ratio: 2.500 Gaps: 3
 Percent Similarity: 52.632 Percent Identity: 38.158

alignment_block:
 US-09-135-238B-2_COPY_273_390 x V88273/rev ..
 Align seg 1/1 to reverse of: V88273 from: 1 to: 266

6 ATGArgLysAlaLeuSerArgArgAlaArgAlaArgLeuAlaValArgMetAr 22
 ||| ||||| ||| |||||
 264 CGAGCTCCCGCGCTCAGCGCTCGCGAGCGCGCTC..... 229

22 gAlaLeuGluSerSerGlnArgProArgGlySerProArgSerG 39
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 228CCTCGCGGGTCCCTCGCGCGCGCGCGCGCC 201

39 InAsnAsnIleTyrSerAlaCysProArgArgAlaArgGlyAlaAspAla 55
 : : : : : ||| ||| : : : : :
 200 CGCGCAGCTCTCCAGAGGCGATAGCG.....AGCGCTGCGCGCTCC 160

56 AlAgLyThrGlyAlaProValProGlyProGlyAlaProLeuProPr 72
 ||| ||| ||| : : : : : ||||| ||||| ||||| |||||
 159 CTAGAGCTGCGACCTCTCCCTGCTCCAGCGCGCGCGCACC....TCCAA 113

72 oAlaProLeuGlnValSerGluSerPro 81
 ||||| : : : : : : : : : : : : : : :
 112 CGCCCGGCTCTCTCCTACCCAAAGCGCC 85

seq_name: N_Geneseq_36:Q10389

seq_documentation_block:
 ID Q10389 standard; DNA; 1903 BP.
 AC Q10389;
 DT 11-APR-1991 (first entry)
 DE Encodes human Retinoic Acid Receptor-gamma A.
 KW human; retinoic acid receptor; hRAR-gamma; embryogenesis; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 415..1779
 FT /*tag= a
 FT /product= human RAR-gamma A
 FT EP-411323-A.
 FT 06-FEB-1991.
 FT 29-JUN-1990; 112469.
 FT 30-JUN-1989; US-374690.
 FT 29-MAR-1990; US-502140.
 FT (INRM) INSERM INST NAT SANTE.
 FT (SQU) LES LABS SQUIBB SA.
 FT Chambon P, Kastner P, Krust A, Petkovich M, Zelent A;
 FT Leroy P, Mendelsohn C, Staub A;
 FT Wp1: 91-038271/06.
 FT P-PSDB; R10550.
 FT Novel human and mouse retinoic acid receptors - encode proteins
 FT used to assay for agonists and antagonists
 FT Claim 8; Fig 5; 33pp; English.
 FT Mouse RAR-gamma cDNA was used to clone its human counterpart from
 FT a T47D breast cancer cell CDNA library. Using a transient
 FT transfection assay in HeLa cells and a reporter gene harbouring a
 FT synthetic responsive element (RE), it was demonstrated that hRAR-
 FT gamma cDNA encodes an RE-inducible transcriptional transactivator.
 FT hRAR-gamma is the predominant RAR RNA species in human skin,
 FT suggesting that it mediates some of the retinoid effects in this
 FT tissue. The nucleotide sequence of hRAR-gamma A upstream from the
 FT initiating codon of the common ORF (positions 130-414) is 77 per
 FT cent homologous to the corresponding 5'UTR of mRAR-gamma cDNA. The
 FT position of the initiating codon is only tentative and is based on
 FT the overall high homology of hRAR-gamma and mRAR-gamma amino acid
 FT sequences.
 FT See also R10547-8, Q10388 and Q10405-8.
 FT Sequence 1903 BP; 395 A; 607 C; 566 G; 335 T;

alignment_scores:
 Quality: 100.00 Length: 96
 Ratio: 2.041 Gaps: 5
 Percent Similarity: 51.042 Percent Identity: 35.417

alignment_block:
 US-09-135-238B-2_COPY_273_390 x Q10389 ..
 Align seg 1/1 to: Q10389 from: 1 to: 1903

20 ArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerProArgPr 36
 ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
 34 CGGTACGTGGGGGGGGGCTCCGAGAGGCCGCCGGGACAGTCCAGC. 81

36 oArgSerGlnAsnAsnIleTyrSerAlaCysPro..... 47
 : : : : : ||| ||| : : : : :
 82CTAGAGCGCTGCCCGCGCGCGCGCGCCAGGCCCCCAG 112

48 .ArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlaProVal 63
 ||||| ||||| : : : : : ||||| ||||| ||||| |||||
 113 TAGCGGAGCGCGGACATTCGAGCGCTCCATCAGAGACTGCGCGAGC 162

64 ProGlyPro GlyAlaProLeuProAlaProLeuGlnValSerGluS 80
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 163 CCGGAGACTCGGGGCTCCCGCGCTCCCTCCCTCC..... 199

80 erProTrpLeuHisAlaProSerLeuLeuThrSerCysGluTyrVal... 95
 ||| ||||| ||||| : : : : : ||||| ||||| ||||| |||||
 200 ..CCACTCCAGTACGGCGCCAGTTCCCTCAAGCTGACCCAGTAGTAGAA 247

96 ...SerLeuTyrHisGlnProAlaAlaMetMetGlu 106
 ||||| : : : : : ||||| ||||| ||||| ||||| ||||| |||||
 248 GCCAGTCTCTCGAGCGCGCGAGCGGACTTTTGAG 283

seq_name: N_Geneseq_36:T05196

seq_documentation_block:
 ID T05196 standard; cDNA; 1903 BP.
 AC T05196;
 DT 17-JUN-1996 (first entry)
 DE Human retinoic acid receptor gamma cDNA.
 KW Retinoic acid receptor; human; RAR; RAR-gamma; transcription factor;
 KW skin; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 415..1779
 FT /*tag= a
 FT /product= retinoic acid receptor gamma protein
 FT misc_feature 415..597
 FT /*tag= b
 FT /*tag= "hRAR-gamma A region"
 FT /note= 598..681
 FT misc_feature 598..681
 FT /*tag= c
 FT /note= "hRAR-gamma B region"
 FT /tag= d
 FT /note= "hRAR-gamma C region"
 FT /tag= e
 FT /note= "hRAR-gamma D region"
 FT /tag= f
 FT /note= "hRAR-gamma E region"
 FT /tag= g
 FT /note= "hRAR-gamma F region"
 FT EP-683227-A1.
 FT 22-NOV-1995.
 FT 29-JUN-1990; 110924.
 FT 30-JUN-1989; US-374690.
 FT 29-MAR-1990; US-502140.
 FT PR

```

22 ArgAlaLeuGluSerSerGlnArgProArglySerProArg..... 35
   ::::| | | | | | | | | | | | | | | | | | | | | |
1020 AAGCGCGCCCTCGCCAGGGGCCCCCTTAATGACACCCACCCCGCG 971
   36 . ProArgSerGlnAsnAsn..... IleTyrSerAlaCysProA 48
   | | | | | | | | | | | | | | | | | | | | | |
970 CCCCCCGCGGAGAACACACCGCTCGCACCTACTTCCACAGCTCA 921
   48 rgrArgAlaArgGlyAlaAspAlaIaGlyThrGlyAlaPro... Val 63
   | | | | | | | | | | | | | | | | | | | | | |
920 GCCGCTGATCTCTTCCAGCGCGGAGCCCTCCCGCCTTCTCTA 871
   64 ProGlyProGlyAlaProLeuProProAla..... ProLeuGly 76
   | | | | | | | | | | | | | | | | | | | | | |
870 CCGCGCGCGCGCGCGGCTTCCCGCCCTTGAAGACCTGCGCTTCA 821
   76 n..... ValSerGluSerProTyrPleuHisAla 86
   | | | | | | | | | | | | | | | | | | | | | |
820 GCCAGCGCGCGCGACTGACTGCTCCGCGCAGCTTGGCTCAGCGCG 771
   86 roserleu 88
   | | | | | | | | | | | | | | | | | | | | | |
770 CAGCGCTG 763
seq_name: N_Geneseq_36:T03650

```

```

seq_documentation_block:
ID T03650 standard; DNA: 1150 BP.
AC T03650;
DC 16-AUG-1996 (first entry)
DE 31-O-desmethyl-FK506 O:methyltransferase gene FKMT2.
KW Streptomycetes; 31-O-desmethyl-FK506 O:methyltransferase; terminator;
   expression cassette; strong promoter; Perm*; ribosome binding site;
   translation start site; transcription terminator; selection marker;
   hygromycin; integration element; multicircle; chromosome; E. coli;
   replicon; ColEI; overexpression; immunosuppressant drug; ss.
OS Streptomyces sp. strain MA6548.
FH Key Location/Qualifiers
   cds 62..844
   /tag= a
   /product= 31-O-desmethyl-FK506 O:methyltransferase
   repeat_unit 842..893
   /tag= b
   /note= "inverted repeat unit"
   repeat_unit 845..877
   /tag= c
   /note= "inverted repeat unit"
FT FT
PN MO9600282-A1.
PD 04-JAN-1996.
PF 19-JUN-1995; U07784.
PR 23-JUN-1994; US-264861.
PA (MERI ) MERCK & CO INC.
PI Motamedi H, Shafiee A;
DR WPI: 96-068866/07
PT Integrative and replicative expression vectors for Streptomycetes
   esp. for expressing 31-O-desmethyl-FK506 O:methyltransferase for
   prodn of the immunosuppressive drug FK506
PS Claim 23; Page 36; 48pp; English.
CC This is the nucleotide sequence of the Streptomycetes gene FKMT2 encoding
   31-O-desmethyl-FK506 O:methyltransferase. The sequence contains the
   coding region for the protein and includes the termination sequence.
   The gene was used in the construction of a Streptomycetes expression
   cassette comprising: (a) a strong promoter e.g. the Perm* constitutive
   promoter, (b) a synthetic ribosome binding site, (c) a translation start
   site (contained within the restriction enzyme site for NdeI), (d) a
   multiple cloning site, (e) a transcription terminator downstream from the
   MCS, esp. the terminator region of the FKMT2 gene. The vector also
   contains a selection marker, pref. the hygromycin resistance gene, an
   integration element, pref. from a multicircle, for integration into the
   Streptomycetes chromosome and an E.coli replicon, pref. the ColEI replicon.
   The vectors can be used for the integration of single copies of desired
   CC genes into the Streptomycetes genome or for the transfer of multiple copies

```

CC of genes of interest for overexpression purposes, e.g. for the conversion
 CC of 31-desmethyl-FK506 to the immunosuppressant drug FK506 by integration
 CC of the FKMT2 gene.
 SQ Sequence 1150 BP; 164 A; 418 C; 396 G; 172 T;

alignment_scores:
 Quality: 96.00 Length: 137
 Ratio: 1.333 Gaps: 6
 Percent Similarity: 52.555 Percent Identity: 29.927

alignment_block:
 US-09-135-238b-2_COPY_273_390 x T03650 ..

Align seg 1/1 to: T03650 from: 1 to: 1150

```

1 LysArgAlaValAlGluArgGlyAlaLeuSerArgAlaArgAla 17
   ::::| | | | | | | | | | | | | | | | | | | | | |
697 CGTCGGGAGGTCCACAGACGTGAGGGCGCGCTCGACAGAGTCCGCG 746
   17 uAlaVal..... ArgMetArgAlaLeuGluSerSerGlnA 29
   | | | | | | | | | | | | | | | | | | | | | |
747 TGCTGCGCGCGCCATGCTTCAACCGCTGTCGCCGAGCAGATCCGCTTC 796
   29 rgrProArglySerProArgProArgSerGlnAsnAsnIleTyrSerAla 45
   | | | | | | | | | | | | | | | | | | | | | |
797 GCCGCGACGACATCCACAGTCCGCC..... ACGGCG 831
   46 CysProArgArgAlaArgGlyAlaAspAlaIaGlyThrGly..... 59
   | | | | | | | | | | | | | | | | | | | | | |
832 TCGGACCCACTGAGCGCGCGCGCGCTACCCGACCGCGCGTCCGG 881
   60 ..... GlnAlaProValPro..... GlyProGlyAla 69
   | | | | | | | | | | | | | | | | | | | | | |
882 TCGCGGCGCTCGCCGACGTCGCGCGCTCTTCCGACCTGCTGCGGCC 931
   69 roLeuProProAlaProLeuGlnValSerGluSerProTyrPleuHisAla 85
   | | | | | | | | | | | | | | | | | | | | | |
932 CTTCACCCGCCAGCTTGGCGGAACAGTCTGTAGTGTCTCCACCTG 981
   86 ProSerLeuLysThrSerCysGluTyrValSerLeuTyrHisGlnPro.. 101
   | | | | | | | | | | | | | | | | | | | | | |
982 CTGCGCGTGAAGACAGCTGCTGGCGATCTCCGTTCGACGCCGAC 1031
   102 ..... AlaAlaMetMetGluAspSerAspSer... AspAspTyrIleA 115
   | | | | | | | | | | | | | | | | | | | | | |
1032 CGCGGCCAGCGCGGCCACCGCGCTCCACATTCGTCAGTACGGGATCC 1081
   115 snValProAla 118
   | | | | | | | | | | | | | | | | | | | | | |
1082 GCTGCCCCGAC 1092
seq_name: N_Geneseq_36:V16346

```

```

seq_documentation_block:
ID V16346 standard; CDNA: 1974 BP.
AC V16346;
DC 03-JUN-1998 (first entry)
DE Nucleotide sequence of human netrin cDNA including 5' and 3' UTRs.
KW Human; netrin; hnMT; ATP binding cassette transporter; hABC3;
   ribosomal L3; RPL3L; augmeneter of liver regeneration; hLR; treatment;
   KW trapping; modulation; expression; antibody; identification; binding;
   KW chemottractant; spinal commissural axon; neural regeneration;
   OS Homo sapiens.
PN MO9748797-A1.
PD 24-DEC-1997.
PF 16-JAN-1997; U00785.
PR 09-DEC-1996; US-762500.
PR 17-JUN-1996; US-665259.
PR 01-OCT-1996; US-720614.
PA (GENZ ) GENZYME CORP.
PI Burn TC, Connors TD, Dackowski WR, Klingner KW, Landes GW,

```

PI Van Raay TJ;
 WPI: 98-063138/06.
 PT Human chromosome 16 genes encoding netrin, ATP binding cassette
 PT transporter, ribosomal L3 and augmentor of liver regeneration
 PT proteins - useful for, e.g. treatment of liver disease and cystic
 PT fibrosis
 PS Claim 5: Fig 4C-D: 220pp: English.
 CC The present sequence represents the nucleotide sequence of human netrin
 CC (hNET) cDNA, including both the 5' and 3' untranslated regions (UTRs).
 CC Netrins define a family of chemotropic factors which have been shown to
 CC play a central role in axon guidance. GRALL2 analysis predicts 6 exons
 CC within the genomic DNA sequence, with 5 exons encoding sequences with
 CC homology to chicken netrins. Chicken netrins have been shown to function
 CC as chemoattractants for developing spinal commissural axons. Human
 CC netrins may therefore have a significant role in neural regeneration.
 CC Though netrins do not by themselves promote axon growth, they do play a
 CC role in the orientation of axon growth. The sequence was isolated using
 CC an exon trap. Sequences encoding human ATP binding cassette transporter
 CC (hABC3), human ribosomal L3 (RpL3L), and human augmentor of liver
 CC regeneration (hALR) were also isolated. The antisense oligonucleotides of
 CC the present sequence are used to modulate expression of hNET prevent its
 CC translation. Antibodies against hNET can be used to block binding of its
 CC naturally occurring ligands. Host cells containing vectors with DNA
 CC inserts encoding the protein can be used in a method for identifying
 CC compounds which bind to hNET.
 S0 Sequence 1974 BP; 252 A; 740 C; 679 G; 303 T;

alignment_scores:
 Quality: 96.00 Length: 161
 Ratio: 1.412 Gaps: 9
 Percent Similarity: 42.236 Percent Identity: 26.708

alignment_block:
 US-09-135-238b-2_COPY_273_390 x V16346/rev ..

Align seg 1/1 to reverse of: V16346 from: 1 to: 1974

```

3 AlalaGlaArgArgAlaLeuSer.....ArgArgAl 14
  ||||| ||||| ||||| ||||| |||||
1972 GCCTTGCCAGCCGCGGCGCTCAGCCGCGCCGCGCTTC 1923
  ||||| ||||| ||||| ||||| |||||
14 aaArgArgLeuAlaValArgMetArgAlaLeuGluSerSer..... 27
  ||||| ||||| ||||| ||||| |||||
1922 GCCTGCTGCAAGCTTCGCGAGCGCGGCTCCATCCAGGGTA 1873
  ||||| ||||| ||||| ||||| |||||
27 ..... 27
1872 GCACGAGGCTTCGCGGCGGCGGATGAGCCCGGCGCCCGCGCG 1823
  ||||| ||||| ||||| ||||| |||||
28 ..... GlnArgPr 30
  ||||| ||||| ||||| ||||| |||||
1822 CCCCAGCGCGGCGCTCAGGCGCGCCCGCCAGAGAGATGAGCGCGCC 1773
  ||||| ||||| ||||| ||||| |||||
30 O.....ArgGlySerProArgProArgSerGlnAsnAlaIleTyrS 44
  ||||| ||||| ||||| ||||| |||||
1772 GGGGAGCAGGCGCGGAGCCGCGAGCCGCGATCCCGCGGCGGCA 1723
  ||||| ||||| ||||| ||||| |||||
44 exAla...CysProArgArgAlaArgGlyAlaAspAlaIleTyrGly 59
  ||||| ||||| ||||| ||||| |||||
1722 GCGCGGTAAGTCCGCGGCGGCGCGCTCTCTGCTCGGAGAACGCGCG 1673
  ||||| ||||| ||||| ||||| |||||
60 .....GlnAlaProValProGlyProGlyAla 68
  ||||| ||||| ||||| ||||| |||||
1672 AGCAGCGCCAGCGGAGAGCGTGCAGCGCGCGCGCTCGCGCGCGC 1623
  ||||| ||||| ||||| ||||| |||||
68 aProLeu..ProProAlaProLeuGlnValSerGluSerProTyrPheIleS 84
  ||||| ||||| ||||| ||||| |||||
1622 ACCCAGCGCGCAGCTGCACGCG.....CAT 1600
  ||||| ||||| ||||| ||||| |||||
85 AlaProSerLeuIleTyr.....SerCys.....G1 93
  ||||| ||||| ||||| ||||| |||||
1599 AGTCCTCTGCAAGACTTCTTAGGCTGATGCGGTAGCTGCGACGAGGGA 1550
  ||||| ||||| ||||| ||||| |||||

```

93 uTyRValSerLeuTyrHisGlnProAlaAla 103
 : ||||| ||||| ||||| ||||| |||||
 1549 GGTTCAGATGTCGAGTCAAGATCGTCGCGGCT 1519

seq_name: N_Geneseq_36:T85537

seq_documentation_block:
 ID T85537 standard; CDNA; 1743 BP.
 AC T85537.
 DT 12-NOV-1997 (first entry)
 DE hNET cDNA sequence from p1 clone 53.8B.
 KW Human; netrin; ATPase binding cassette transporter; ribosomal L3;
 KW augmentor of liver regeneration; hNET; hABC3; SEM L3; hALR;
 KW chromosome 16; exon trapping; axon; chicken; laminin domain;
 KW C. elegans; UNC-6; ss.
 OS Homo sapiens.
 PN WC9702346-A2.
 PD 23-JUN-1997.
 PF 17-JUN-1996; U10469.
 PR 30-JUN-1995; US-000596.
 PA (GENZ) GENZYME CORP.
 PI Burn TC, Connors TD, Dackowski WR, Klinger KW, Landes GM;
 PI Van Raay TJ;
 DR WPI: 97-108959/10.
 DR P-PDB: W25670.
 PT New isolated human chromosome 16 genes - encode netrin, ATPase
 PT binding cassette transporter, ribosomal L3 sub-type or augmentor of
 PT liver regeneration
 PS Claim 3: Fig 4; 98pp: English.
 CC This sequence encodes the human netrin protein (hNET). The genomic
 CC sequence was isolated from human chromosome 16 by exon trapping.
 CC Netrins are a family of diffusible factors involved in axon growth.
 CC hNET cDNA is 66% homologous to the chicken netrin-2 cDNA. The protein
 CC sequence has 58.6% identity and 73.4% similarity to chicken netrin-2.
 CC The conserved N-terminal laminin domains found between chicken netrin-1
 CC and -2, C. elegans UNC-6 and the murine laminin protein B2 are also
 CC conserved in hNET. Domain V contains 3 epidermal growth factor repeats
 CC and is 81.1% identical between hNET and chicken netrin-2. Domain C has
 CC the characteristic basic residue bias with an isoelectric point of 12.26
 CC and has a partial RGD domain for potential adhesion/signaling receptors.
 CC The hNET gene can be used to develop chemoattractants for use in
 CC axon regeneration.
 S0 Sequence 1743 BP; 231 A; 652 C; 589 G; 271 T;

alignment_scores:
 Quality: 95.50 Length: 147
 Ratio: 1.540 Gaps: 8
 Percent Similarity: 42.177 Percent Identity: 27.211

alignment_block:
 US-09-135-238b-2_COPY_273_390 x T85537/rev ..

Align seg 1/1 to reverse of: T85537 from: 1 to: 1743

```

12 ArgArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSer.. 27
  ||||| ||||| ||||| ||||| |||||
1721 CCGCGTTCGCGCTCGCTGCAAGCTTCGCGAGCGCGCGTCCAGCGTCCG 1672
  ||||| ||||| ||||| ||||| |||||
27 ..... 27
1671 CAGGGTACACAGAGGCTTCGCGGCGGCGGATGAGCCCGGCGCCGCGG 1622
  ||||| ||||| ||||| ||||| |||||
27 ..... 27
1621 CCCCAGCGCGCCAGCGCGGCTCCAGGCGCGCCCGCCAGAGAGGTAG 1572
  ||||| ||||| ||||| ||||| |||||
28 GlnArgPro.....ArgGlySerProArgSerGlnAsnAs 41
  ||||| ||||| ||||| ||||| |||||
1571 CCGCGCGCGGAGAGCAGCGCGCGAGCCGCGATCCCGCGGCGG 1522
  ||||| ||||| ||||| ||||| |||||
41 nIleTyrSerAla...CysProArgArgAlaArgGlyAlaAspAlaIle 57
  ||||| ||||| ||||| ||||| |||||

```



```

      : :::::::::::::: ::::::::::::::::::::
1521 CACCACAGCGCTACTCCGCGCGGCGCTCTCTCCGCTCCGGA 1472
      : :::::::::::::: ::::::::::::::::::::
57 1yThrgLy.....GluAlaProValProGly 65
      : :::::::::::::: ::::::::::::::::::::
1471 ACACGGGAGACCGCCACCGGAGAGGTGTCACGCGCGCGGCGCTCG 1422
      : :::::::::::::: ::::::::::::::::::::
66 ProGlyAlaProLeu.ProProAlaProLeuGlnValSerGluSerProt 82
      : :::::::::::::: ::::::::::::::::::::
1421 CCGCGCGACCCACCGCCACCTGCACCG..... 1394
      : :::::::::::::: ::::::::::::::::::::
82 rPLeuHISAlaProSerLeuLysThr.....SerCys 92
      : :::::::::::::: ::::::::::::::::::::
1393 .....CATAGTCTCTTGCAGAACTTTTAGCGTATGCGGTAAGTGC 1349
      : :::::::::::::: ::::::::::::::::::::
93 .....GluTyValSerLeuTyHisGlnProAlaAla 103
      : :::::::::::::: ::::::::::::::::::::
1348 CACGGGACAGTTTGACGTGCGAGTCACAGTCTGTGGGCT 1310
      : :::::::::::::: ::::::::::::::::::::

seq_name: N_Geneseq_36.V16306
seq_documentation_block:
ID V16306 standard; cDNA; 1743 BP.
AC V16306:
DE 03-JUN-1998 (first entry)
DE cDNA encoding human netrin.
KW Human; netrin; hNET; ATP binding cassette transporter; hABC3;
KW ribosomal L3; RPL3L; augmentor of liver regeneration; hALR; treatment;
KW trapping; modulation; expression; antibody; identification; binding;
KW chemottractant; spinal commissural axon; neural regeneration;
KW axon growth; orientation; substrate specificity; ligand; exon trap; ss.
OS Homo sapiens.
FH Key
FT 1. 1743
FT CDS /product= netrin

FN W09748797-A1.
PD 24-DEC-1997.
PE 16-JAN-1997; U00785.
PR 09-DEC-1996; US-762500.
PR 17-JUN-1996; US-665259.
PR 01-OCT-1996; US-720614.
PA (GENZ ) GENZYME CORP.
PI Burn TC, Connors TD, Dackowski WR, Klingner KW, Landes GM,
PI Van Raay TJ;
PI WPI; 98-063138/06.
DR P-PSDB; W46758.
PT Human chromosome 16 genes encoding netrin, ATP binding cassette
PT transporter, ribosomal L3 and augmentor of liver regeneration
PT proteins - useful for, e.g. treatment of liver disease and cystic
PT fibrosis
PS Claim 4; Fig 4A-B; 220bp; English.
CC The present sequence encodes human netrin (hNET). Netrins define a family
CC of chemotropic factors which have been shown to play a central role in
CC axon guidance. GAIT2 analysis predicts 6 exons within the genomic DNA
CC sequence, with 5 exons encoding sequences with homology to chicken
CC netrins. Chicken netrins have been shown to function as chemottractants
CC for developing spinal commissural axons. Human netrins may therefore have
CC a significant role in neural regeneration. Though netrins do not by
CC themselves promote axon growth, they do play a role in the orientation of
CC axon growth. The sequence was isolated using an exon trap. Sequences
CC encoding human ATP binding cassette transporter (hABC3), human ribosomal
CC L3 (RPL3L), and human augmentor of liver regeneration (hALR) were also
CC isolated. The antisense oligonucleotides of the present sequence are used
CC to modulate expression of hNET prevent its translation. Antibodies
CC against hNET can be used to block binding of its naturally occurring
CC ligands. Host cells containing vectors with DNA inserts encoding the
CC protein can be used in a method for identifying compounds which bind
CC to hNET.
SQ Sequence 1743 BP; 231 A; 652 C; 589 G; 271 T;
alignment_scores:

```

```

      Quality: 95.50      Length: 147
      Ratio: 1.540      Gaps: 8
Percent Similarity: 42.177      Percent Identity: 27.211

alignment_block:
us-09-135-238b-2_copy_273_390 x V16306/rev ..
Align seg 1/1 to reverse of: V16306 from: 1 to: 1743

12 ArgAlaAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSer.. 27
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1721 CGCGTTCGCTGCTGTCAGAGCTCCGACAGCGCGCGCTCCACGCGTCC 1672
27 .....
27 .....
1671 CCAGGTAGACAGAGGCTTCGCGGCGCGGAGTAGAGCCGGGCCCCGGC 1622
27 .....
1621 CCGCGGCGCCCGACGCGCGGCTCCAGGCCCGCCCCCAGAGAGTAG 1572
28 GlnArgPro.....ArgGlySerProArgProArgSerGlnAsnAs 41
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1571 CGCGCGCGCGGAGACAGCGCGGCGAGCCGCGCGCATCCCGCGCGG 1522
41 nileTySerAla...CysProArgArgAlaArgGlyAlaAspAlaAlaG 57
: :::::::::::::: ::::::::::::::::::::
1521 CACCACAGCGCGCTACTCCGCGCGGCGCTCTCTCCGCTCCGGA 1472
      : :::::::::::::: ::::::::::::::::::::
57 1yThrgLy.....GluAlaProValProGly 65
      : :::::::::::::: ::::::::::::::::::::
1471 ACACGGGAGACCGCCACCGGAGAGGTGTCACGCGCGCGGCGCTCG 1422
      : :::::::::::::: ::::::::::::::::::::
66 ProGlyAlaProLeu.ProProAlaProLeuGlnValSerGluSerProt 82
      : :::::::::::::: ::::::::::::::::::::
1421 CCGCGCGACCCACCGCCACCTGCACCG..... 1394
      : :::::::::::::: ::::::::::::::::::::
82 rPLeuHISAlaProSerLeuLysThr.....SerCys 92
      : :::::::::::::: ::::::::::::::::::::
1393 .....CATAGTCTCTTGCAGAACTTTTAGCGTATGCGGTAAGTGC 1349
      : :::::::::::::: ::::::::::::::::::::
93 .....GluTyValSerLeuTyHisGlnProAlaAla 103
      : :::::::::::::: ::::::::::::::::::::
1348 CACGGGACAGTTTGACGTGCGAGTCACAGTCTGTGGGCT 1310
      : :::::::::::::: ::::::::::::::::::::

seq_name: N_Geneseq_36.T93095
seq_documentation_block:
ID T93095 standard; cDNA; 24379 BP.
AC T93095:
DE 11-MAY-1998 (first entry)
DE Streptomyces frenolicin gene cluster.
KW Frenolicin; antibiotic; feed additive; anticoccidial;
KW coccidiostatic; efflux pump; butyrate starter synthase;
KW polypeptide synthase; PKS; hemiketalase; ketoreductase; cyclase;
KW dehydratase; ketoreductase; hydroxylase; Streptomyces roseofulvus;
KW ds.
OS Streptomyces sp.
FH Key
FT 636..2948
FT CDS /tag= a
FT /product= 80 kDa non-membrane protein
FT /note= "gene A (specifically claimed)"
FT 2945..3916
FT CDS /tag= b
FT /product= membrane protein
FT /note= "gene B (specifically claimed)"
FT 4020..4844
FT CDS /tag= c
FT /product= protein with 6 membrane-spanning domains
FT /note= "gene C (specifically claimed)"
FT 4841..6415
FT CDS /tag= d

```

FT /product= ATP-binding component of ABC transporter
FT /note="gene D (specifically claimed)"
FT CDS
FT /tag= e
FT /product= unknown non-membrane protein
FT /note="gene E (specifically claimed)"
FT CDS
FT /tag= f
FT /product= putative efflux pump
FT /note="gene F (specifically claimed)"
FT CDS
FT /tag= g
FT /product= transcription activator
FT /note="gene G (specifically claimed)"
FT CDS
FT /tag= h
FT /product= translationally coupled to gene I
FT /note="gene H (specifically claimed)"
FT CDS
FT /tag= i
FT /product= homologue of fabH
FT /note="gene I (specifically claimed)"
FT CDS
FT /tag= j
FT /product= acyl carrier protein
FT /note="gene J (specifically claimed)"
FT CDS
FT /tag= k
FT /product= acyltransferase
FT /note="gene K (specifically claimed)"
FT CDS
FT /tag= l
FT /product= PKS ketosyl synthase subunit
FT /note="gene L (specifically claimed)"
FT CDS
FT /tag= m
FT /product= PKS chain length factor
FT /note="gene M (specifically claimed)"
FT CDS
FT /tag= n
FT /product= acyl carrier protein used by the PKS
FT /note="gene N (specifically claimed)"
FT CDS
FT /tag= o
FT /product= putative hemiketal dehydrase
FT /note="gene O (specifically claimed)"
FT CDS
FT /tag= p
FT /product= ketoreductase related to actIII
FT /note="gene P (specifically claimed)"
FT CDS
FT /tag= q
FT /product= cyclase/dehydrase related to act VII
FT /note="gene Q (specifically claimed)"
FT CDS
FT /tag= r
FT /product= cyclase/dehydrase related to act V
FT /note="gene R (specifically claimed)"
FT CDS
FT /tag= s
FT /product= oxidoreductase
FT /note="gene S (specifically claimed)"
FT CDS
FT /tag= t
FT /product= quinone-forming hydroxylase
FT /note="gene T (specifically claimed)"
FT CDS
FT /tag= u
FT /product= not specified
FT /note="gene U"
EP-806480-A2.
PD 12-NOV-1997.
PF 02-MAY-1997; 107329.

PR 07-MAY-1996; US-016753.
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
PI Reeves CD, Soliday CL;
DR WPI: 97-538619/50.
DR P-PDB: W34199-W34219.
PT Streptomycetes frenolicin gene cluster - useful for producing
PT recombinant frenolicin antibiotics
PS Claim 1; Page 40-60; 66pp; English.
CC This DNA sequence comprises the Streptomycetes frenolicin gene
CC cluster containing specifically claimed coding sequences (genes
CC A-U) that respectively encode 21 proteins (see W34199-219) involved
CC in frenolicin synthesis. The genes can be divided into 5
CC subclusters: (1) genes A, B, C, D and F encode an efflux pump; (2)
CC genes H, I, J and K encode butyrate starter synthases; (3) genes
CC L, M and N encode polyketide synthases (PKS); (4) genes O, P, Q and
CC R encode a hemiketalase, a ketoreductase and cyclases/dehydrases;
CC and (5) genes S and T encode a keto/enoyl reductase and a
CC hydroxylase. Also claimed are vectors, host cells (especially a
CC Streptomycetes sp., particularly Streptomycetes roseofulvus), and the
CC encoded proteins. Cells transformed using the above sequence can
CC be cultured to produce frenolicins or frenolicin precursors. The
CC precursors can be converted to frenolicins by chemical or other
CC methods. The frenolicins can be oxidised to frenolicin B, an
CC antibiotic used as an anticoccidial agent. The frenolicins can be
CC used as animal feed additives.
SQ Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T;

alignment_scores:
Quality: 95.00 Length: 135
Ratio: 1.439 Gaps: 9
Percent Similarity: 48.889 Percent Identity: 32.593

alignment block:

US-09-135-238B-2_COPY_273_390 x T93095 ..

Align seg 1/1 to: T93095 from: 1 to: 24379

2 ATGAAValAlGluArgLysAlaLeuSerArgArgAlaArgLys 18
||||| ||||| |||
21737 CGGCGAGTACGCGCGGTACCGCGCGGTACGCGAGAGAGCGCGC 21786
18 aValArgMetArgAlaLeuGluSer.....SerLArgProArgLys 33
..... |||||
21787 CACCGGCGCGAGTGCACCGCTCCGAGCGCAGAGCGCGTCCGCGAG 21836
33 eProArg.....ProArgSerGlnAsn 41
||||| ||||| |||||
21837 TTCGCGAAGCGCGCTTCGCGCGCGCGCGCGCGCGCGCAAGTGC 21886
42 IleArgSer.....AlaCysProArgArgAla..... 50
||| ||||| |||||
21887 GGGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGACC 21936
51ArgGlyAlaAspAlaAlaGlyThrGlyLys 61
21937 CCGCTCTCTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAT 21986
61 IapProVal.....ProGlyProGlyAlaProLeuPro... 71
||||| ||||| ||||| |||||
21987 CCGCGCGCGGTTCGCGAGCGGTCCGAGG...GGCGTCCGCGCGCGT 22033
72 ProAlaProLeuGlnValSerGluSerPro.....TripleuHi 84
||||| ||||| ||||| |||||
22034 CCGCGCGCGCGGTCCGCGCGGTTCGCGCGCGCGCGAGAGGCGTGCAT 22083
84 sAlaProSerLeuLysThrSerCysGluLysValSerLeuLysGlnP 101
||||| ||||| ||||| |||||
22084 TGCCTCGTCA.....TCGCTAGCTACCTCCACGCGAACAATCTCC 22124
101 roAla 102
|||||
22125 CGGCG 22129

```
seq_name: N_Geneseq_36.V25925
seq_documentation block:
ID      V25925 standard; cDNA; 24379 BP.
AC      15-JUL-1998 (first entry)
DE      Streptomyces roseofulvus frenolicin gene cluster.
KW      Streptomyces roseofulvus; frenolicin gene cluster; frenolicin B,
        antibiotic; ss.
OS      Streptomyces roseofulvus.
        location/Qualifiers
FH      Key
        CDS
FT      /*tag= a
FT      /transl_except= (pos:636..638,aa:Met)
FT      /note= "encodes protein given in W55800"
FT      2945..3916
FT      CDS
FT      /*tag= b
FT      /transl_except= (pos:2945..2947,aa:Met)
FT      /note= "encodes protein given in W55801"
FT      4020..4844
FT      CDS
FT      /*tag= c
FT      /transl_except= (pos:4020..4022,aa:Met)
FT      /note= "encodes protein given in W55802"
FT      4841..6415
FT      CDS
FT      /*tag= d
FT      /note= "encodes protein given in W55803"
FT      6533..7183
FT      CDS
FT      /*tag= e
FT      /transl_except= (pos:6533..6535,aa:Met)
FT      /note= "encodes protein given in W55804"
FT      7344..8897
FT      CDS
FT      /*tag= f
FT      /note= "encodes protein given in W55805"
FT      9164..10012
FT      CDS
FT      /*tag= g
FT      /transl_except= (pos:9164..9166,aa:Met)
FT      /note= "encodes protein given in W55806"
FT      complement (10105..10621)
FT      CDS
FT      /*tag= h
FT      /transl_except= (pos:10621..10619,aa:Met)
FT      /note= "encodes protein given in W55807"
FT      complement (10618..11628)
FT      CDS
FT      /*tag= i
FT      /transl_except= (pos:11628..11626,aa:Met)
FT      /note= "encodes protein given in W55808"
FT      11809..12066
FT      CDS
FT      /*tag= j
FT      /note= "encodes protein given in W55809"
FT      complement (12154..13209)
FT      CDS
FT      /*tag= k
FT      /transl_except= (pos:13209..13207,aa:Met)
FT      /note= "encodes protein given in W55810"
FT      13409..14686
FT      CDS
FT      /*tag= l
FT      /transl_except= (pos:13409..13411,aa:Met)
FT      /note= "encodes protein given in W55811"
FT      14767..16047
FT      CDS
FT      /*tag= m
FT      /note= "encodes protein given in W55812"
FT      16120..16371
FT      CDS
FT      /*tag= n
FT      /note= "encodes protein given in W55813"
FT      complement (16453..16935)
FT      CDS
FT      /*tag= o
FT      /transl_except= (pos:16935..16933,aa:Met)
FT      /note= "encodes protein given in W55814"
FT      17088..18903
FT      CDS
FT      /*tag= p
FT      /note= "encodes protein given in W55815"
FT      17903..18898
FT      CDS
FT      /*tag= q
FT      /note= "encodes protein given in W55816"
FT
```

```

FT      CDS
FT      18893..19839
FT      /tag= r
FT      /note= "encodes protein given in W55817"
FT      complement (19990..20907)
FT      CDS
FT      /tag= s
FT      /transl_except= (pos:20907..20905,aa:Met)
FT      /note= "encodes protein given in W55818"
FT      CDS
FT      complement (20904..22094)
FT      /tag= t
FT      /transl_except= (pos:22040..22038,aa:Arg)
FT      /transl_except= (pos:22037..22035,aa:Arg)
FT      /transl_except= (pos:22034..22032,aa:Thr)
FT      /transl_except= (pos:22031..22029,aa:lys)
FT      /note= "encodes protein given in W55819"
FT      PD
FT      J10094395-A.
FT      PD
FT      14-APR-1998.
FT      PD
FT      07-MAY-1997. 116652.
FT      PR
FT      04-APR-1997. US-042935.
FT      PR
FT      07-MAY-1996. US-016753.
FT      PA
FT      (HOFF.) HOFFMANN LA ROCHE & CO AG F.
FT      WP1: 98-279231/25.
FT      DR
FT      P-PSDB; W55800-W55819.
FT      PT
FT      firenolicin gene cluster - useful as an antibiotic
FT      PS
FT      Claim 2; Page 36-44; 50pp; Japanese.
FT      CC
FT      The present sequence represents a firenolicin gene cluster from
FT      streptomyces roseofulvus. The present invention describes: (1) a
FT      CC
FT      vector containing the firenolicin gene cluster DNA sequence operably
FT      CC
FT      connected to an expression control sequence; (2) a host cell
FT      CC
FT      transformed by the above vector; (3) a protein coded by the above DNA
FT      CC
FT      sequence; (4) a method for the preparation of firenolicin or a
FT      CC
FT      biosynthetic intermediate for it in which the above cell is cultured
FT      CC
FT      and firenolicin or its biosynthetic intermediate is isolated from the
FT      CC
FT      culture or the cell; (5) a method for the preparation of firenolicin B
FT      CC
FT      by oxidising firenolicin, and (6) a method for the preparation of a
FT      CC
FT      feed composition by mixing firenolicin with other components.
FT      CC
FT      firenolicin B is useful as an antibiotic.
FT      SQ
FT      Sequence 24379 bp; 3077 A; 9792 C; 8499 G; 3011 T;

```

```

alignment_scores:
  Quality: 95.00      Length: 135
  Ratio: 1.439      Gaps: 9
Percent Similarity: 48.889      Percent Identity: 32.593

alignment block:
US-09-135-238B-2_COPY_273_390 x V25925 ..

Align seq 1/1 to: V25925 from: 1 to: 24379

2 ArgAlaValaIguArgrArgIysAlaLeuSerArgArgAlaAargArgLeuAl 18
|||||  |||||  |||  |||||  |||
21737 CGGGCAGAGTACGCGGGCGGTACCGCCGCGCTAGCGAGAGCGAGCCGCC 21786

18 aVaIaArgMeIaArgAlaLeuGluSer.....serGlnaArgProArgGlyS 33
|||||  |||||  |||  |||||  |||
21787 CAGCGCGCCAGAGTGTGAGGCCGTCGCCGAGGCCAGAGACGCCCTCGGGCAG 21836

33 ePrOArg.....ProArgserGlnaasn 41
|||||  |||||  |||||  |||||  |||
21837 TTCGCCGAGACCCGCTTCGAGCGCCGCGCTGGGCGCGGAGCAGAAAGTGCC 21886

42 IletYser.....AlaCysProArgArgAla..... 50
|||  |||||  |||||  |||||  |||
21887 GGGCGAGAGCCCGCGCCGACAGATCCCGCGAGAGACTTCCCGCGAGACC 21936

51 .....ArgIlaIaaspAlaIaagIyThrIyGluA 61
|||||  |||||  |||||  |||||  |||
21937 CCGGTCTCTCTACCGCGCGCGCCCTGCGCGCGCGAGACCGCGCGAGCG 21986

61 IapProVal.....ProGlyIProGlyAlaProLeuPro... 71
|||||  |||||  |||||  |||||  |||
21987 CCCCAGGCCCTTCGCGAGCGAGTCCGGGG...GGCTCTCCCGGCCCTCTCT 22033

```


Percent Similarity: 50.000 Percent Identity: 34.375

alignment_block:
US-09-135-238b-2_copy_273_390 x T05197 ..

Align seg 1/1 to: T05197 from: 1 to: 637

```
20 ArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerProArgPr 36 *
   ||| |||:::|||||:::|||||:::|||||:::|||||
34 CGGTCACGAGTGGGCGGGCGGCTCCGAGAGGCCCGGGACAGTCCAGC.. 81
   ||| |||:::|||||:::|||||:::|||||
36 oArgSerGlnAsnAsnIleTyrSerAlaCysPro..... 47
   ||| |||:::|||||:::|||||
82 .....CTAGAGCGGTGCCCCCGGAGAGCCCCCAG 112
   ||| |||:::|||||:::|||||
48 ..ArgArgAlaArgGlyAlaAspAlaIleGlyThrGlyAlaProVal 63
   ||| |||:::|||||:::|||||
113 TACGGCGACCCCGGACATTCGACGCTCCATCCAGCAGCTGCGCGAGC 162
   ||| |||:::|||||:::|||||
64 ProGlyPro.GlyAlaProLeuProProAlaProLeuGlnValSerGlu 80
   ||| |||:::|||||:::|||||
163 CCGGGACCTCGGGGCTCCGCGCTCCCTCCCTC..... 199
   ||| |||:::|||||:::|||||
80 ePrProTPrLeuH1AlaProSerLeuTyrThrSerCysGlyTyrVal... 95
   ||| |||:::|||||:::|||||
200 ..CCACTCCAGCTACGGCCAGTTCCTCCATCACTGACCGCATGTATGAA 247
   ||| |||:::|||||:::|||||
96 ...SerLeuTyrHisGlnProAlaAlaMetMetGlu 106
   ||| |||:::|||||:::|||||
248 GCCAGTCTCTGACGCGCGACGCGGACTTTTGGAG 283
```

seq_name: N_Geneseq_36:T78508

seq_documentation_block:

ID T78508 standard; DNA; 44377 BP.

AC T78508;

DE 26-FEB-1998 (first entry)

DE platenolide synthase gene cluster.

KW platenolide synthase gene cluster; platenolide production; smtg gene;

KW multi-functional protein; macrolide antibiotic; spiramycin; ss.

OS Streptomyces ambofaciens.

OS location/Qualifiers

FT CDS 350..14002

FT FT /*tag- a /transl_except- (pos:350..352, aa:Met)

FT FT /note- "ORF1 encodes protein shown in W23716"

FT CDS 14046..20036

FT FT /*tag- b /note- "ORF2 encodes protein shown in W23717"

FT CDS 20110..31284

FT FT /*tag- c /transl_except- (pos:20111..20113, aa:Met)

FT FT /note- "ORF3 encodes protein shown in W23718"

FT CDS 31329..36071

FT FT /*tag- d /note- "ORF4 encodes protein shown in W23719"

FT CDS 36155..41830

FT FT /*tag- e /note- "ORF5 encodes protein shown in W23720"

FT EP-791656-A2.

PD 27-AUG-1997.

PF 19-FEB-1997; 301066.

PF 22-FEB-1996; US-012050.

PA (ELIL) LILLY & CO ELI.

PI Burgett SG, Kunstloss SA, Rao RN, Richardson MA;

PI Rostock PR;

DR WPI; 97-418047/39.

DR P-PSDB; W23716-W23720.

PT for production of streptomycetes ambofaciens platenolide synthase domain -

PS Claim 9; Pages 8-33; 81pp; English.

CC This sequence represents the platenolide synthase gene cluster of the

CC invention. This sequence is referred to as the smtg gene, and was

CC isolated from Streptomyces ambofaciens. This sequence encodes the
CC multi-functional proteins which direct the synthesis of the polyketide
CC platenolide. Platenolide is the basic building block of the macrolide
CC antibiotic spiramycin. The DNA can be used to produce compounds
CC exhibiting antibiotic activity based on the platenolide structure,
CC including specifically the macrolide antibiotic spiramycin and spiramycin
CC analogues and derivatives. Modifications of the platenolide synthase DNA
CC sequence can be made so as to change the number and type of carboxylic
CC acids incorporated into the growing polyketide chain and to change the
CC kind of post-condensation processing that is conducted.
S0 Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T;

alignment_scores: Quality: 94.00 Length: 159
Ratio: 1.593 Gaps: 5
Percent Similarity: 37.107 Percent Identity: 24.528

alignment_block:
US-09-135-238b-2_copy_273_390 x T78508/rev ..

Align seg 1/1 to reverse of: T78508 from: 1 to: 44377

```
2 ArgAlaValAlaGluArgTyrGlyAlaLeuSerArgArgAlaArg..... 15
   |||::: |||::: |||::: |||
28270 AGGACTCGATCTCGGGACCGGCGCTCCCTCGCGGCGACCTGATG 28221
   |||::: |||::: |||::: |||
16 .....ArgLeuAlaValArgM 21
   ||| |||::: |||
28220 CGGACCGGCTCTCGGAGAGCGGGCGAGACGCGCACGTGCAGCGGTGCTG 28171
   ||| |||::: |||
21 eArgAlaLeuGluSerSerGlnArgPro.....Arg 31
   ||| |||::: |||
28170 TCCGGCGCGGTACAGCCGACGACCGCGCGGCGGAGCGCATCGGAGGC 28121
   ||| |||::: |||
32 GlySerProArgProArgSerGlnAsnAsnIleTyrSerAlaCysProAr 48
   ||| |||::: |||
28120 GGGCGTCTCTCGGGAGAGAGCCCGGAGGCTCATGTCTGACGCCCGCG 28071
   ||| |||::: |||
48 gArgAla..... 50
   ||| |||::: |||
28070 TCCAGGACGCGGGATGATGCCGAAGCGGCGCGCTCGCGGCGCTG 28021
   ||| |||::: |||
51 .....ArgGlyAlaAspAlaIleGlyThrGlyGlu 60
   ||| |||::: |||
28020 GAGGCGCAGCGGACCTCCGCGGTGATGTCGGGCGCGCAGCGGTGGCGC 27971
   ||| |||::: |||
61 AlaProValProGly..... 65
   ||| |||::: |||
27970 TCACCAAGTCCCGGAGAAAGCGGCGGTAGCCGTAAACGAGTTCGGAGAG 27921
   ||| |||::: |||
66 .....ProGlyAlaProL 70
   ||| |||::: |||
27920 CGGTGTAGAAAGTCCCGATGTCACAGTTGTCGGCGCGGGGGCGGGGT 27871
   ||| |||::: |||
70 euProPro.....Ala 73
   ||| |||::: |||
27870 GGGCGCGAGCGGTCCGGCGCGTCCGCGCGGTCTCGCGGTGAGCAGCC 27821
   ||| |||::: |||
74 ProLeuGlnValSerGluSerProTPr 82
   ||| |||::: |||
27820 CCTCGGCTGACAGGATCCAGTCCGTGG 27794
```

a

...

[illegible]

5'UTR
sig_peptide
CDS

```

/codon_start=1
/product="binds and transports polymeric immunoglobulin"
/function="polymeric immunoglobulin receptor"
/protein_id="AAA67440.1"
/db_xref="GI:458422"

```

```

/translation="MRLEFTLELVTFEGSVSTSPIFGEOEVSSLEGDSVSLICIP
TSVNRHTRKYWCROGASGMCITLIISSNGYLSKESGRANLINPEPNTIVINIEOLT
DDTGSKYCGLGTSNRLGSLFDVSLSEVSOYPELPSLTHVYKNDIGRNTIDCPKRENV

```


ICDAGEPSAKKNDVLOVLAPELLELYKDRSSVTEPCDLGRENANKEATYICRMKE
TCDVINTLGRKRDPEFERILLITPKDNGRNSVITLTKREDAGHOCGSHSGLEJOE
GMPIDTQWLFVNESTIPNRNSVAVGTGSAVALACINPKNKSSSLKTCRMWGDENG
HCPALVGTQAOVQOEYEGRLAFDQNGTYYTLNLTEDAGFYWCLNNGSRMT
TLELOVATREPNLEVTYVONATVILEGTETVYLPCKEYSEKYNKSNKCHIL
PSHDEGARSSVSCDOSQSLVNTINVSKEDEMYCYKOCQGTGETATYIAVEE
RTGSSHNPTDANAKRAKVALEEVYSSISEKEMKIPNPGFANRETONVQDAO
ENRASGASGADGOSRSSSVLFTVPLGIAVALIYVAVRHRKNDNRMSIS
SYKRDISADDFKNSRDGNDNGASDPDTQYIEGDELYTTTECTAPEBESKAKR
SSKEADMAISAFLOSTIAOVHDPQEA

mat_peptide 139.2397
misc_feature /note="ectoplasmic region, binds polymeric immunoglobulin"
misc_feature /note="transmembrane segment of the receptor"
misc_feature /note="cytoplasmic tail of the receptor"
3'UTR /note="cytoplasmic tail of the receptor"
BASE COUNT 861 a 796 c 784 g 654 t
ORIGIN

alignment_scores:
Quality: 198.00 Length: 168
Ratio: 1.980 Gaps: 5
Percent Similarity: 59.524 Percent Identity: 31.548

alignment_block:
US-09-135-238B-2 x MMU06431 ..

Align seg 1/1 to: MMU06431 from: 1 to: 3095

10 PheupProValSerGlyAlaLeuArgIleuProGluValLysValG1 26
121 TTTTACGGGGCTCCACAAAAGCCCATTTGGTCCCGAGAGATAG 170
26 uGlyGluLeuGlySerValThrIleLysCysProLeuProGlu... 41
171 TAGATAGAGAGCGGCTCTTCATCAGTCACTACGACAGACACCT 220
42MetHisValArgIleTyrLeuCysArgIleMetAlaGlySer 55
221 CTGTCAACGGCGCACACCCGAAATACGTGTCGCGAAGAGACC...AGC 267
56 GlyThrCysGlyThrValValSerThrThrAsnPhelLeuAlaGlyTyr 72
268 GGCATGTGCACAAAGCGCTCATCTCAATGCTGCTCCTCCAGAGACTA 317
72 rLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeu 89
318 TTGAGGACAGACCACTCATCACTCCAGAGAACACATTGTGA 367
89 alGluValThrGlnLeuThrIleuSerAspSerGlyValTyrAlaCysGly 105
368 TTAACATTGACGACCTCACCGACGACGACGAGCTACAGAGTGGC 417
106 AlAGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVa 122
418 CTGGGT...ACCAGTAAACGAGCGCTGCTTCGATGACACGAGAGT 464
122 lHisSerGluTyrGluProSerTyrPGLuGlnProMetProGluThrP 139
465 C...AGCCAGGTCTCTTAAGTTCG 484
139 rLysTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSer 155
485 CGAGTGACACCCACGTC.....TACACAAG 510
156 SerSerLysPheValThrArgValThrThrProAlaGlnArgGlyLysVa 172
511 GACATAGGACGAATGTGACCATGTGATGCCCTTTCAAAAGGAGAAATGT 560
172 lPro 173

561 TCCC 564

seq_name: gb_om:AF091137

seq_documentation_block: 2848 bp mRNA MAM 01-JUL-1999

LOCUS AF091137 2848 bp mRNA MAM 01-JUL-1999
DEFINITION Trichosurus vulpecula polymeric-immunoglobulin receptor precursor
(PIGR) mRNA, complete cds.

ACCESSION AF091137
VERSION AF091137.1 GI:5305512

KEYWORDS Trichosurus vulpecula.

SOURCE Trichosurus vulpecula.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Metatheria; Diprotodontia; Phalangeridae; Trichosurus.

REFERENCE 1 (bases 1 to 2848)
AUTHORS Adamski, F.M. and Demmer, J.
TITLE Cloning and characterization of PIGR and J chain of the marsupial,
Trichosurus vulpecula (brush-tailed possum)

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2848)

AUTHORS Adamski, F.M. and Demmer, J.
TITLE Direct Submission
SUBMITTED (09-SEP-1998) Dairy Science, AgResearch Ruakura, East
Street, Hamilton, New Zealand

FEATURES
location/Qualifiers
1..2848
/organism="Trichosurus vulpecula"
/db_xref="taxon:9337"
/tissue_type="mammary gland"
1..2848
/gene="PIGR"
149..2350
/gene="PIGR"
/codon_start=1
/product="polymeric-immunoglobulin receptor precursor"
/protein_id="AAD41688.1"
/db_xref="GI:5305513"
/translation="MAFLACLLALLPVYKSPKIFGKQVTEGEGSVSIQCFYSS
SVRHRKXFLONROSCETIVSSNGVSEFSGRAFLTFPNNSFLIOISLEKE
DILYCGKFGTSLNROSLSPDITLVEGNGVSEFSGRAFLTFPNNSFLIOISLEKE
RKIKCKKDESSALVDSOYGPYTRARALISGSSRPVYVTSQIRQVGVY
CGVEDSDTGIONNDKLEPEPELAEAGSTTLCAAGSTVAVSPFELCONAK
ETCDLVINSGFTNNATGRILFSTFTGSPSLITOVREDCGVHCAODEGDP
EECPRIALDLFVSEETVPSPLVYKSGSGSVITCYDKRNNILKYCKNKGSS
CTRLVSLGVDESEYEGRYALWDEPENGIFVYILNLPDAGVYMLNSGENRKS
VKRIENDGOLPLIAPRYAOLGSLTISCHFPCKEYSERYCKNSNOGCTLPQ
EGSSOAFVDCNONSRSVLTLSYVRDHEGYMGVYNGVNGVYCTIVSVSEEVSG
NAIOPNATLNDAYPRYKREIEVPTDLSSTEHSGSSVYTVPLATVYGA
VALGIITAKRRMFSORVSVGSTRDLSSELENRPQGANENKASVQETTLGGEDE
IITQENPEIIOQPKOVKSSKEEDMAVYTAFLLOSNNANETPLDDPSNA"

gene
CDS

BASE COUNT 769 a 769 c 722 g 667 t
ORIGIN

alignment_scores:
Quality: 192.00 Length: 204
Ratio: 1.684 Gaps: 8
Percent Similarity: 55.882 Percent Identity: 30.392

alignment_block:
US-09-135-238B-2 x AF091137 ..

Align seg 1/1 to: AF091137 from: 1 to: 2848

17 LeuArgIleLeuProGluVal..... 23
173 CTGTGCTCTCCCTCTGCTGTTTCATAGAAAGCCCATATTGGCCCA 222
24 LysValGluGlyLeuGlySerValThrIleLysCysProLeuP 40
223 GCAAGTGACAGAGAGTGAAGAGATCTCTCATCAGAGTCTTACC 272


```
117:  gb_gss7: *
118:  gb_gss8: *
119:  gb_gss9: *
120:  gb_gss10: *
121:  gb_gss11: *
122:  gb_gss12: *
123:  gb_gss13: *
124:  gb_gss14: *
```

Result	Query				
No.	Score	Match	length	DB	ID
					Description

C 1	635.8	33.3	711	44	A1760373	A1760373	wh87d11.
C 2	516.4	27.0	578	35	A10848226	A10848226	oy72g09.1
C 3	458.8	24.0	524	79	AA4552045	AA4552045	UI-H-B13
C 4	431	22.6	509	44	A1760350	A1760350	wh87b03.3
C 5	431	22.6	511	44	A1763709	A1763709	wh93d08.3
C 6	418.2	21.9	456	39	A13997232	A13997232	th28g08.3
C 7	418.2	21.9	488	30	AA766222	AA766222	th28g09.3
C 8	409.2	21.4	453	30	AA769441	AA769441	nz41f03.5
C 9	403.2	21.1	441	73	AA0001912	AA0001912	wh81g03.3
C 10	400	20.9	458	44	A1761021	A1761021	wh10b11.3
C 11	399	20.9	476	26	AA4565309	AA4565309	z133h01.5
C 12	390.8	20.5	474	30	AA703531	AA703531	z133h01.5
C 13	382.4	20.0	410	79	AA4001870	AA4001870	UI-HF-BKOC
C 14	380	19.9	404	26	AA456707	AA456707	aa13g04.
C 15	374	19.6	528	90	W01276	W01276	2a40f12.r1
C 16	360.2	18.9	418	40	A1475802	A1475802	tc94f09.3
C 17	359	18.8	444	35	A1141006	A1141006	oy68f06.3
C 18	357.4	18.7	411	30	AA744586	AA744586	ny25g04.8
C 19	353	18.5	441	31	AA806272	AA806272	oc26a03.5
C 20	348.8	18.3	390	27	AA502725	AA502725	ng63f12.5
C 21	337	17.6	416	80	AA512632	AA512632	xx77c09.3
C 22	332	17.4	414	39	A1434224	A1434224	l33g406.3
C 23	329.2	17.2	367	33	AA991282	AA991282	os81a12.5
C 24	328.6	17.2	368	40	A1475832	A1475832	tc87a11.3
C 25	327.4	17.1	417	31	AA813368	AA813368	nz08a02.8
C 26	325.4	17.0	430	91	W86442	W86442	2b61a12.s1
C 27	321.8	16.8	419	44	A1813566	A1813566	wh18f08.x
C 28	318.6	16.7	365	35	A1324090	A1324090	q13d07.x
C 29	318	16.6	398	39	A1395553	A1395553	tc90c06.x
C 30	316.2	16.6	335	39	A1433629	A1433629	l33f104.x
C 31	314.2	16.5	352	81	AA571597	AA571597	xx33g10.x
C 32	311.2	16.3	414	32	AA8683421	AA8683421	nm25a04.s
C 33	311.2	16.3	430	31	AA832318	AA832318	oc95h05.s
C 34	305	16.0	387	39	A1439627	A1439627	tc91i06.x
C 35	303.4	15.9	312	79	AA402953	AA402953	UI-HF-BKOC
C 36	297.8	15.6	340	38	A1783751	A1783751	td40h02.x
C 37	295	15.4	340	78	AA737573	AA737573	xx33h01.x
C 38	287.2	15.0	518	82	AA6555210	AA6555210	1058b3
C 39	273.2	14.3	341	35	A1081685	A1081685	oy71g07.x
C 40	266.2	13.9	322	63	A18655238	A18655238	wh80h10.x
C 41	265.2	13.9	382	34	A1057449	A1057449	oc90c02.x
C 42	259	13.6	494	82	AA655218	AA655218	105873
C 43	257.2	13.5	369	31	AA832322	AA832322	oc95h11.s
C 44	254	13.3	287	31	A1083368	A1083368	oy71g06.x
C 45	254	13.3	482	79	AA444691	AA444691	UI-H-B13

RESULT 1
AI760373/C

AI760373 711 bp mRNA EST 20-DEC-1999
 wh87d11.x1 NCI_CGAP_CLL1 Homo sapiens CDNA clone IMAGE:2387733
 similar to TR:060667 060667 ANTI-FAS-INDUCED APOPTOSIS. ;, mRNA

ACCESSION	sequence.
VERSION	A1760373
KEYWORDS	A1760373.1 GI:5176040
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
JOURNAL	NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap .
COMMENT	1 (bases 1 to 711)
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
	Tumor Gene Index
	Unpublished (1997)
	On Oct 30, 1998 this sequence version replaced gi:3812483.

Unpublished (1997)
On Oct 30, 1998 this sequence version replaced gi:3812483.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
M.D., Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLN at:
www.bio.llnl.gov/brfp/image/image.html
Insert length: 1449 Std Error: 0.00
Seq primer: -40up from Gibco
High quality sequence stop: 472.
Location/Qualifiers
1..711

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1 IMAGE:2387733"
/clone_1b="NCI-CCAP-CLL1"
/tissue_type="B-cell, chronic lymphocytic leukemia"
/lab_host="DH10B"
/notes="(vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dn) primer [5',
TGTTTCCATCTGGAAGTGGAGGCGCGCCGATCTTTTTTTTTTTTTTTTTTTT
T 3'] : double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pRTT3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

```

[illegible]

OM of: US-09-135-238B-2 to: N_Geneseq_36:* out_format : pfs
Date: Sep 12, 2000 6:55 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet_p2n_model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09135238/runat_24082000_091324_17918/app_query.fasta_1.1282
-DB=N_Geneseq_36 -QPMT=fastap -SUFFIX=P2n.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -DOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cd1 -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=100000
-USER=US09135238_@cgn1_1_123 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT_THREADS=1

Information block:

US-09-135-238B-2
Query length: 390
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 121.040000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:X28178	+	2047.00	2129.34	5.3e-111	Human PIGRL-1 coding sequence.
N_Geneseq_36:X28179	+	1125.00	1174.09	8.5e-58	Human PIGRL-1 coding sequence
N_Geneseq_36:X281290	+	198.00	199.21	0.0017	Mouse poly-immunoglobulin rece
N_Geneseq_36:X281291	+	188.00	170.27	0.0697	Mouse poly Ig receptor protei
N_Geneseq_36:X281292	+	187.50	192.24	0.0042	cDNA for human immunity relat
N_Geneseq_36:X281293	+	185.50	185.74	0.0096	Rat poly-immunoglobulin rece
N_Geneseq_36:X281294	+	182.00	189.53	0.0059	Secretory component DNA fragme
N_Geneseq_36:X281295	+	182.00	186.82	0.0083	Human secretory immunoglobulin
N_Geneseq_36:X281296	+	182.00	186.01	0.0093	Secretory component DNA fragme
N_Geneseq_36:X281297	+	182.00	183.02	0.0136	Human poly-immunoglobulin rece
N_Geneseq_36:X281298	+	182.00	181.23	0.0171	Bovine poly-immunoglobulin rec
N_Geneseq_36:X281299	+	171.00	174.33	0.0414	Rat kidney injury related mole
N_Geneseq_36:X281300	+	171.00	172.62	0.0515	Rat kidney injury related mole
N_Geneseq_36:X281301	+	166.50	170.51	0.0676	Clone 1 for truncated poly Ig
N_Geneseq_36:X281302	+	166.50	165.34	0.1312	Rabbit poly-immunoglobulin rec
N_Geneseq_36:X281303	+	159.50	167.34	0.1015	Human PIGR-2 coding sequence.
N_Geneseq_36:X281304	+	145.00	148.11	1.20	Clone 2 for truncated poly Ig
N_Geneseq_36:X281305	+	143.50	133.93	7.37	Plasmid pD17-hm14.H1. Inhibi
N_Geneseq_36:X281306	+	142.00	130.90	0.8361	Human EST R2810 DNA. Complexes
N_Geneseq_36:X281307	+	139.50	130.12	12.02	Plasmid expressing HBR96-A. In
N_Geneseq_36:X281308	+	138.50	146.63	1.44	M. tuberculosis immunogenic pol
N_Geneseq_36:X281309	+	136.50	149.12	1.05	EST clone EN260. New polynucle
N_Geneseq_36:X281310	+	136.50	140.14	3.32	Human dUTPase gene 5' region.
N_Geneseq_36:X281311	+	136.00	123.79	27.06	Plasmid pAH4602. Antibody con
N_Geneseq_36:X281312	+	135.00	123.35	28.60	Plasmid pAH4625. Antibody con
N_Geneseq_36:X281313	+	131.00	118.16	55.71	Plasmid pAH4807. Antibody con
N_Geneseq_36:X281314	+	130.00	133.70	7.59	Antibody D heavy chain. Prodn.
N_Geneseq_36:X281315	+	129.50	134.18	7.13	Primate anti-human B7.1 ant
N_Geneseq_36:X281316	+	129.50	134.18	7.13	Macaque primatized 7C10 heavy
N_Geneseq_36:X281317	+	127.00	128.05	15.56	TS1096 transposon. Leucine aux
N_Geneseq_36:X281318	+	126.50	120.88	39.26	Plasmid pHC65 nucleotide sequ
N_Geneseq_36:X281319	+	126.50	114.03	94.63	Approximate nucleotide sequen
N_Geneseq_36:X281320	+	125.50	129.22	13.48	Human anti-HBs heavy chain. Hu
N_Geneseq_36:X281321	+	125.50	122.89	30.35	3F4 human G2/64 chimeric antib
N_Geneseq_36:X281322	+	125.50	116.64	67.64	Human calcium channel a1B subu
N_Geneseq_36:X281323	+	124.50	116.45	69.37	Anti-IgE VH expression vector
N_Geneseq_36:X281324	+	124.50	108.70	187.39	HSV-2 strain SB5 Contig ID 12
N_Geneseq_36:X281325	+	124.50	108.70	187.39	acidovorans gamma-lactanase
N_Geneseq_36:X281326	+	124.00	124.32	25.27	CD4-specific CDR-grafted heavy

N_Geneseq_36:T50962	+	124.00	115.30	80.34	7073	TFB-509 CDR-grafted heavy c
N_Geneseq_36:V57472	+	123.50	125.47	21.80	1929	Sorghum bicolor (L.) Moench
N_Geneseq_36:T96035	+	123.00	125.54	21.60	1795	Human kidney injury related
N_Geneseq_36:T72685	+	123.00	113.19	105.27	8051	Human biosynthesis gene clu
N_Geneseq_36:X25774	+	123.00	113.08	106.78	8160	S. erythraea erythromycin-sy

seq_name: N_Geneseq_36:X28178

seq_documentation_block:

ID X28178 standard; cDNA: 2040 BP.
AC X28178;
DT 16-JUN-1999 (first entry)
DE Human PIGRL-1 coding sequence.
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;
KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;
KW diagnosis; therapy: ss.
OS Homo sapiens.
PN EP-905238-A2.
PD 31-MAR-1999.
PF 14-AUG-1998; 306487.
PR 30-OCT-1997; US-961564.
PR 25-AUG-1997; US-056935.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Sweet RW, Truneh A, Wu S;
DR WPI: 99-192666/17.
DR P-PSDB: Y05001.
PT New polypeptides encoding PIGRL-1 useful for treating diseases such
as X-linked Severe Combined Immunodeficiency
PS Claim 4; Page 18-19; 26pp; English.
CC This sequence encodes the human PIGRL-1 protein of the invention.
CC Autoimmune diseases involving altered expression or activity of PIGRL-1
may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined
Immunodeficiency (XSCID) and IgA deficiency. These diseases can be
diagnosed or susceptibility to them predicted by: (1) determining whether
there is a mutation in the genomic copy of the gene encoding PIGRL-1; or
(2) measuring the amount of PIGRL-1 in a sample derived from the patient.
CC Patients deficient in PIGRL-1 can be treated by administering either the
PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.
CC Patients with excessive expression or activity of PIGRL-1 can be treated
by administering an antagonist of PIGRL-1, an antisense nucleic acid
molecule which inhibits the expression of PIGRL-1 or administering
sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can
be used to identify its agonists by contacting a cell expressing PIGRL-1
with a candidate compound in the presence of a signal system and noting
the candidate as an agonist if a signal is produced. The same method can
be used to identify antagonists of PIGRL-1 but the presence of an
antagonist is indicated by a decrease in production of the signal.
CC Antibodies against PIGRL-1 may be used to isolate or identify clones
expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to
identify chromosomal mutations in the gene encoding PIGRL-1 in patients.
CC This information may then be correlated with the incidence of autoimmune
disease in those patients to identify whether the mutation causes the
disease.

Sequence 2040 BP; 478 A; 616 C; 494 G; 452 T;

alignment_scores:

Quality: 2047.00 Length: 390
Ratio: 5.252 Gaps: 0
Percent Similarity: 99.744 Percent Identity: 99.744

alignment_block:

US-09-135-238B-2 x X28178

Align seg 1/1 to: X28178 from: 1 to: 2040

1 MetAspArgTrpLeuTrpProLeuTyPheLeuProvalserGlyAlaLe 17
||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
146 ATGGACTTCTGGCTTGGCCACTTTACTTCTGCCAGTATCAGGGCCCT 195
17 uArgileLeuProGluValIysValGluGlyLeuGlyGlySerValT 34
||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
196 GAGGATCTCTCCAGAGTAAGGTAGAGGGGAGCTGGGGGGATCAGTTA 245

34 hrleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 50
246 CCATCAAGTCCGCACCTCTCCGAAATGATGTGAGGATATATCTGTGCGG 295
51 GluMetAlaGlySerGlyThrCysGlyThrValSerThrThrAspNph 67
296 GAGATGCTGATCTGGAACATCTGTACCTGTTGATCCACCAACTT 345
67 eileLysAlaGlyThrGlyArgValThrLeuLysGlnThrProArgL 84
346 CATCAAGGAGAAATCAAGGGCGGAGTTACTCTGAAGCAATACCCAGCA 395
84 yAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
396 AGAATCTGTTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGGA 445
101 ValThrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117
446 GTCTATGCTTGGGAGGCGGATGAACACAGACCGGGGAAAGACCCAGAA 495
17 sValThrLeuAsnValHisSerGlyThrGluProSerTrpGluGlnP 134
496 AGTCACCTGAATGTCCACAGTGAATACGAGCATCATGGGAGAGCAGC 545
134 roMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGluMet 150
546 CAATGCTGAGACTCAAAATGGTTTCATCTGCTTCTTCTTCCAGATG 595
151 ProAlaTyrAlaSerSerSerLysPheValThrArgValThrProAl 167
596 CCTGCATATGCCAGTCTCTCCAAATTCGTAAACAGAGTTACACACAGC 645
167 aGlnArgGlyLysValProProValHisHisSerSerProThrThrGlnI 184
646 TCAAGGGGCAAGTCCCTCCAGTTCACACCTCTCCCTCCACACCCAAA 695
184 leThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLys 200
696 TCACCCACCCCTCCAGTGTCCAGAGCATCTTCAGTAGCAGGTGACAAG 745
201 ProArgThrPheLeuProSerThrThrAlaSerLysLysLeuAlaLeuG 217
746 CCCCAGAACCTTCTGCTGCTTCCACTACAGCTCCTCAAAATCTCAGTCTGGA 795
217 uGlyLeuLeuLysProGlnThrProSerTyrAsnHisHisThrArgLeuH 234
796 GGGGCTCTCAAGCCGACAGCCAGCCAGTACAAACCCACACACAGGCTGC 845
234 lsArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGln 250
846 ACAGCCAGAGACACTGGAGTATGCTCTACAGTCTGGGAGGAGAGCCAA 895
251 GlyPheHisLeuLeuProThrLeuGlyLeuPheLeuAlaLe 267
896 GGATTTTCATCTGATCCGACCATCTCTGGGCTTTTCTCTGCTGGCACT 945
267 uLeuGlyLeuValValTyrArgAlaValClnArgArgLysAlaLeuSerA 284
946 TCTGGGCTGCTGCTGTAAGGGCGGCTTCAAGAGGAGGAAAGCCCTCTCCA 995
284 rArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln 300
996 GCGGGGCGCCGACATGGCGCTGAGGAGTGGCGCTGGAGAGCTCCAG 1045
301 ArgProArgGlySerProArgProArgSerGlnAsnAsnIleTyrSerAl 317
1046 AGGCCCCGCGGCTGCGGCGGCGGCTCCCAAAACACATCTACAGCGC 1095
317 acYsProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlap 334
1096 CTGCCCGCGCGCTCGTGGAGGAGGAGCTGACAGGACAGGGAGGCC 1145

334 roValProGlyProGlyAlaProLeuProProAlaProLeuGlnValSer 350
1146 CCGTTCCCGCCGCGGAGCGGTGGCCCCCGCCGCGCTGCGAGTGTCT 1195
351 GluSerProTrpLeuHisAlaProSerLeuLysThrSerCysGlyTrpVa 367
1196 GAATCTCCCTGGCTCATGCCCATCTCTGAAGACACAGCTGTGAATACGT 1245
367 lSerLeuTyrHisGlnProAlaAlaMetMetGluAspSerAspSerAspA 384
1246 GAGCCTCTACCCAGCGCTGGCCCATGTGAGGACAGTTCAGATG 1295
384 sPyrLeuAsnValProAla 390
1296 ACTACATCAATGTTCTGCTGCC 1315

seq_name: N_Geneseq_36.X28179

seq_documentation_block:

ID X28179 standard; CDNA; 1047 BP.

AC X28179;

DT 16-JUN-1999 (first entry)

DE Human PIGRL-1 coding sequence fragment.

KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;

KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;

KW diagnosis; therapy; ss.

OS Homo sapiens.

PN EP-905238-A2.

PD 31-MAR-1999.

PF 14-AUG-1998; 306487.

PR 30-OCT-1997; US-961564.

PR 25-AUG-1997; US-056935.

PA (SMJK) SMITHKLINE BEECHAM CORP.

PI Sweet RW, Truneh A, Wu S;

DR WPI: 99-192666/17.

DR P-PSDB: Y05002.

PT New polypeptides encoding PIGRL-1 useful for treating diseases such

as X-linked Severe Combined Immunodeficiency

PS Disclosure: Page 8; 26pp; English.

CC This sequence encodes the human PIGRL-1 protein of the invention.

CC Autoimmune diseases involving altered expression or activity of PIGRL-1

CC may include hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined

CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be

CC diagnosed or susceptibility to them predicted by: (1) determining whether

CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or

CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.

CC Patients deficient in PIGRL-1 can be treated by administering either the

CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.

CC Patients with excessive expression or activity of PIGRL-1 can be treated

CC by administering an antagonist of PIGRL-1, an antisense nucleic acid

CC molecule which inhibits the expression of PIGRL-1 or administering

CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can

CC be used to identify its agonists by contacting a cell expressing PIGRL-1

CC with a candidate compound in the presence of a signal system and noting

CC the candidate as an agonist if a signal is produced. The same method can

CC be used to identify antagonists of PIGRL-1 but the presence of an

CC antagonist is indicated by a decrease in production of the signal.

CC Antibodies against PIGRL-1 may be used to isolate or identify clones

CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to

CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.

CC This information may then be correlated with the incidence of autoimmune

CC disease in those patients to identify whether the mutation causes the

CC Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T;

alignment_scores:

Quality: 1125.00

Ratio: 4.429

Percent Similarity: 92.364

Percent Identity: 90.182

alignment_block:

US-09-135-238B-2 x X28179

Length: 275

Gaps: 12

Percent Identity: 90.182

Align seg 1/1 to: X28179 from: 1 to: 1047

```

1 MetAspArgTrpLeuTrpProLeuTyPheLeuProValSerGlyAlaLe 17
150 ATGGACATCTGCTGGCTTTGGCCACATTACTCTCCAGTATCAGGGCCCT 199
17 uArgLeuProGluValLysValGluGlyGluLeuGlyGlySerValT 34
200 GAGGATCTCCCAAGAGTAAGTAGAGGGGGAGCTGGCGGATCAGTTA 249
34 hrLeuLysCysProLeuProGluMetHisValArgIleTyLeuCysArg 50
250 CCATCAAGTCCCACTTCTTGAATGATGTGAGGATATATCTGTCCGG 299
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67
300 GAGATGCTGGATCTGGAACATGTGTACGTGTATCCACCAACAT 349
67 eileLysAlaGluTyLysGlyArgValThrLeuLysGlnTyProArgL 84
350 CATCAAGGCAAGATACAGGGCGGAGTTACTCTGAAGCAATACCCAGCA 399
84 yAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
400 AGAATCTGTCTTAGTGGAGGTAAACAGCTGACAGAAAGTGACAGCGGA 449
101 ValTyAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnTy 117
450 GTCTATGCTGCGGACG. GCATGACACAGACCGGGGAAAGACCCAGAA 498
117 svalThrLeuAsnValHisSerGluTyGluProSerTrpGluGlnP 134
499 AGTCACCTGAATGTCCACAGTAGTAATACGAGCATCATGGGAAGACAG 548
134 roMetProGluThrProLysTrpPheHisLeuProTyLeuPheGlnMet 150
549 CAATGCTGAGACTCCAAATAGTTTCACTGCTGCTTATTTGTCAGATG 598
151 ProAlaTyAlaSerSerLysPheValThr. ArgValThrThrPro 166
599 CCTGCATATCGCGTTCTTCCACATTCGTAAACCGCAGAGTTACACACCA 648
167 Ala.GlnArgGlyLysValProProValHisHisSerSerProThrThrG 183
649 GCTTCAAGGGGCAAGGTCTCCAGTTCCACACTCTCCCGCCACCC 698
183 lnIle.ThrHisArgPro.ArgValSerArgAlaSerValAlaGlyA 199
699 AAATTCACCCACCGCTTCCAGTGTCAGATGTCAGATCTTTCAGTAGCAGGTG 748
199 spLysProArg.ThrPheLeuProSerThrThrAlaSerLysLysSerAl 215
749 ACAAGCCCCGAAACTTCTCTGCCATCCACTACAGCTCAAAATCTCAGC 798
215 aLeuGluGlyLeuLeuLys...ProGlnThrProSerTyThrAsnHisHist 231
799 TCTGGAAGGCTGTGTTCAAGCCCCAGAGCGCCAGCTACAAACAN.CACA 847
231 hrArgLeuHisArgGlnArgAlaLeuAsp...TyLysSerGlnSerG 246
848 CCAGGCTGCACAGGCAGAGAGACTGGATACTTATGGNTACAGTCTGG 897
246 yArgGluGly.....GlnGlyPhe...HisIleLeuIlePro.ThrIle 259
898 G...GAGGGGAANGNCCAAGATTTTTCATTTCTGATTCCTCCCGGACCATC 944
260 LeuGlyLeuPhe 263
945 NTTGGGCTTT 956

```

seq_name: N_Geneseq_36:T31290

seq_documentation_block:

```

ID T31290 standard; cDNA; 3095 BP.
AC T31290;
DT 24-FEB-1997 (first entry)
DE Mouse poly-immunoglobulin receptor, cDNA.
KW Mouse; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; murine; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT cds 85..2400
   /*tag= a
   WO9621012-A1.
   PD 11-JUL-1996.
   PF 27-DEC-1995; U16889.
   PR 30-DEC-1994; US-367395.
   PR 04-MAY-1995; US-434000.
   PA (PLAN-) PLANT BIOTECHNOLOGY INC.
   PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
   PA (PLAN-) PLANET BIOTECHNOLOGY INC.
   PI Hiatt AC, Lehner T, Ma JKC;
   DR WPI: 96-333987/33.
   DR P-PSDB; W03180.
   PT Immunoglobulin and protection protein complex and its prodn. in
   PT plants - useful for passive immunisation against mucosal antigens,
   PT esp. against S. mutans and S. sorbinus to prevent dental caries
   PS Disclosure: pages 117-121; 152pp; English.
   CC The present sequence encodes the mouse poly-immunoglobulin (Ig)
   CC receptor, a portion of which corresp. to residues 1-627, pref.
   CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
   CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
   CC The Ig of the invention comprises a PP as above in association with
   CC an Ig derived heavy chain, having at least a portion of an antigen
   CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
   CC gastrointestinal, environments, therefore enhancing its
   CC effectiveness in passively immunising animals against mucosal
   CC pathogens. The Ag binding domain is specifically derived from the
   CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
   CC by binding, e.g. Streptococcus mutans serotypes C, e and f, or
   CC S. sorbinus serotypes d and g.
   SQ Sequence 3095 BP; 861 A; 796 C; 784 G; 654 T;

```

alignment_scores:

```

Quality: 198.00 Length: 168
Ratio: 1.980 Gaps: 5
Percent Similarity: 59.524 Percent Identity: 31.548

```

alignment_block:

US-09-135-238B-2 x T31290 ..

Align seg 1/1 to: T31290 from: 1 to: 3095

```

10 PheLeuProValSerGlyAlaLeuArgIleLeuProGluValLysValG 26
121 TTTTCAGGGGTCTCCACAAAAGCCCATATTTGGTCCCGCAGGAGTGAG 170
26 uGlyLeuLeuGlyGlySerValThrIleLysCysProLeuProGlu... 41
171 TAGTATAGAGGCGACCTCTGTTTCCATCAGCTGTACTACCCAGACCT 220
42 .....MetHisValArgIleTyLeuCysArgGluMetAlaGlySer 55
221 CTGTCAACCGGCACACCGGAAATACTGGTCCGCAAGGAGCC...AGC 267
56 GlyThrCysGlyThrValValSerThrThrAsnPheIleLysAlaGluTy 72
268 GGCATGTGCACACGCTCATCTCTCAATGGCTACCTCTCCAAGGAGTA 317
72 rLysGlyArgValThrLeuLysGlnTyProArgLysAsnLeuPheLeuV 89

```


LOCUS AW402953 312 bp mRNA 16-FEB-2000
 DEFINITION UI-HF-BK0-abd-c-06-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone
 IMAGE:3055859 5', mRNA sequence.
 ACCESSION AW402953
 VERSION AW402953.1 GI:6921743
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 312)
 AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT On May 18, 1998 this sequence version replaced gi:3138550.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1350
 Email: Robert.Strausberg@nih.gov
 Eco RI site shown at the beginning of the sequence.
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 CDNA Library Preparation: M.B. Soares Lab
 CDNA Library Arrayed by: M.B. Soares Lab
 DNA Sequencing by: M.B. Soares Lab
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
 Seq primer: M3 Forward

FEATURES
 source
 1..312
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3055859"
 /clone_lib="NIH_MGC_36"
 /tissue_type="lymph"
 /cell_type="germinal center B cells"
 /cell_line="MGC85"
 /lab_host="DH10B (LTI)"
 /note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
 Constructed from size fractionated cytoplasmic mRNA
 (0.5-1.5kb). Directionally cloned. Cells provided by Louis
 M. Staudt, Ph.D. Library preparation by Maria de Fatima
 Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
 97 a 74 c 87 g 54 t

alignment_scores:
 Quality: 544.00 Length: 101
 Ratio: 5.386 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-135-238B-2 x AW402953

Align seg 1/1 to: AW402953 from: 1 to: 312
 41 GluMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyth 57
 |||||
 10 GAAATGATGTGAGGATATATCTGTCCCGGAGATGCTGGATCTGGAAC 59
 57 rCysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysG 74
 |||||
 60 ATGTGGTACCGTGGTATCCACACCACTCATCAAGGAGATACAAAG 109
 74 lyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 90
 |||||
 110 GCCAGTACTCTGAACAATATCCACGCAAGAAATCTGTTCTCTAGTGGAG 159
 91 ValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaG 107
 |||||
 160 GTAACACAGCTGACAGAAAGTACACGGAGTCTATGCTCGCGGCGGG 209
 107 yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 124

|||||
 210 CATGACACAGACCGGGAAGACCCAGAAAGTACCCTGATGTCACA 259
 124 erGluTyrGluProSerTrpGluGluGlnProMetProGluThrProLys 140
 |||||
 260 GTGAATAGAGCCATCATGTGGAAGAGCAGCAATGCTGAGACTCCAAA 309
 141 Trp 141
 310 TGG 312
 seq_name: gb_est45:AW655210
 seq_documentation_block:
 LOCUS AW655210 518 bp mRNA EST 05-APR-2000
 DEFINITION 105863 MARC 1B0V Bos taurus cDNA 5', mRNA sequence.
 ACCESSION AW655210
 VERSION AW655210.1 GI:7421036
 KEYWORDS EST.
 SOURCE Bos taurus.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 518)
 AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
 Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
 Keele,J.W.
 TITLE Design and use of four pooled tissue normalized cDNA libraries for
 EST discovery in cattle
 JOURNAL Unpublished (2000)
 COMMENT On Mar 8, 1999 this sequence version replaced gi:4388119.
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCCAGTCAGCAGC
 Plate: 86 row: K column: 10
 Seq primer: ATTAGTGACACTAG.

FEATURES
 source
 1..518
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 1B0V"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from lymph node, ovary,
 fat, hypothalamus, and pituitary."
 BASE COUNT 124 a 133 c 157 g 104 t
 ORIGIN

alignment_scores:
 Quality: 522.00 Length: 136
 Ratio: 4.424 Gaps: 0
 Percent Similarity: 86.765 Percent Identity: 68.382
 alignment_block:
 US-09-135-238B-2 x AW655210
 Align seg 1/1 to: AW655210 from: 1 to: 518
 1 MetAspArgTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLe 17
 |||||
 110 ATGGACCTGTGGCTTGGCGCTTTACTTCTCTCAGTAGTCGAGGCC 159

AP

LOCUS AW402953 312 bp mRNA EST 16-FEB-2000
DEFINITION UI-HF-BK0-abd-c-06-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone IMAGE:3055859 5', mRNA sequence.
ACCESSION AW402953
VERSION AW402953.1 GI:6921743
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 312)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
JOURNAL On May 18, 1998 this sequence version replaced gi:3138550.
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.lnli.gov/bbrp/image/image.html
Seq primer: M13 Forward.
FEATURES
source
1..312
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3055859"
/clone_lib="NIH_MGC.36"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MC85"
/lab_host="DH10B (LTR)"
/note="vector: pT73-Pac; Site_1: NotI; Site_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (0.5-1.5kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT 97 a 74 c 87 g 54 t
ORIGIN

alignment_scores:
Quality: 544.00 Length: 101
Ratio: 5.386 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-35-2388-2 x AW402953
Align seg 1/1 to: AW402953 from: 1 to: 312
41 GluMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThr 57
10 GAATGTCATGTGAGGATATATCTGCGCGGAGATGCTGTGATCTGGAAC 59
57 rCysGlyThrValValSerThrThrAsnPhelIleLysAlaGluTyrLysG 74
60 ATGTGGTACCGTGGTATCCACCACCACTTCATCAGGCAGAACATCAAGG 109
74 lYArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 90
110 GCGGAGTTACTCTGAACAATACCCACCAAGATCTGTTCTCTAGTGGAG 159
91 ValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaG 107
160 GTAACACAGCTGACAGAAAGTGACAGCGAGTCTATGCTCGGAGCGGG 209
107 yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 124

210 CATGAACACAGACCGGGGAAACACCCAGAAAGTCACCTGATGTCCACA 259
124 erGluTyrGluProSerTrpGluGluGlnProMetProGluThrProLys 140
260 GTGAATACAGCCATCATGGGAAGAGCAGCCCAATGCTGAGACTCCAAA 309
141 Trp 141
310 TGG 312
seq_name: gb_est45:AW655210
seq_documentation_block:
LOCUS AW655210 518 bp mRNA EST 05-APR-2000
DEFINITION 105863 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW655210
VERSION AW655210.1 GI:7421036
KEYWORDS EST.
SOURCE Bos taurus.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 518)
AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and Keeler,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT On Mar 8, 1999 this sequence version replaced gi:4388119.
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.
PCR primers
FORWARD: AGGAACACAGCTATGACCAT
BACKWARD: GTTTCACGATCAGCAGC
Plate: 86 row: K column: 10
Seq primer: ATTTAGGTGACACTATAG.
FEATURES
Location/Qualifiers
1..518
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."
BASE COUNT 124 a 133 c 157 g 104 t
ORIGIN

alignment_scores:
Quality: 522.00 Length: 136
Ratio: 4.424 Gaps: 0
Percent Similarity: 86.765 Percent Identity: 68.382
alignment_block:
US-09-135-2388-2 x AW655210
Align seg 1/1 to: AW655210 from: 1 to: 518
1 MetAspArgTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLe 17
110 ATGGACCTGTGGCTTTGGCGCTTTACTTCTCCAGTAGTCGGAGCCCC 159

PN WO9857993-A1.
PD 23-DEC-1998.
PE 10-JUN-1998. U11975
PR 19-JUN-1997; US-050969.
PA (REGC) UNIV CALIFORNIA.
PI Chintalacharuvu KR, Morrison SL;
DR WPI; 99-080950/07.
DR P-PSDB; W95601.
PT Producing secretory immunoglobulin in single cells - useful to
PT produce commercial quantities of secretory immunoglobulin to prevent
PT or treat infections
PS Disclosure; Page 22; 39pp; English.
CC The sequence is that of the coding region for the secretory
CC component of human secretory immunoglobulin A (sIgA).
CC The sequence can be used as part of a method for the
CC production of sig molecules. This method is useful for
CC producing commercial quantities of sig (especially sIgA) to treat
CC or prevent infections. In particular, sIgA produced by the method
CC can be used to prevent or treat infections in mammals, birds or
CC fish; especially systemic infections or infections at a mucosal
CC surface. It is especially useful to prevent or treat infection
CC with human immunodeficiency virus (HIV), respiratory syncytial
CC virus, flu virus or cold virus. The method allows production of
CC commercial quantities of sig molecules for therapeutic use, not
CC previously possible; production using non-plant cells and a
CC single cell type is more efficient than a previous multi-step
CC process of fusing recombinant plant cells, and avoids alterations
CC of the sig by plant cells. sIgA molecules are more stable
CC and resistant to proteolysis than previously used IGA molecules,
CC and can be administered to prevent as well as to treat infections,
CC unlike e.g. IgG and IgM molecules.
SQ Sequence 1839 BP; 423 A; 504 C; 544 G; 368 T;

alignment_scores:

Quality: 180.50 Length: 106
Ratio: 2.314 Gaps: 4
Percent Similarity: 73.585 Percent Identity: 38.679

alignment_block:

US-09-135-238B-2_COPY_18_253 x X07407 ..

Align seg 1/1 to: X07407 from: 1 to: 1839

4 ProGluValLysValGluGlyGluLeuGlySerValThrIleLysCys 20
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
73 CCGGAG...GAGGTGAATAGTGTGAAGGTAACCTCAGTGTCCATCAGCTG 119
| ||| :|||:|||||:|||||:|||||:|||||:|||||
20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33
| ||| :|||:|||||:|||||:|||||:|||||:|||||
120 CTACTACCCACCCACCTCTGTCAACCGGCACACCGGAAGTACTGGTGCC 169
| ||| :|||:|||||:|||||:|||||:|||||:|||||
33 rGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
|:|:| | ||| :|||:|||||:|||||:|||||:|||||: |||
170 GGCAGGGAGCTAGAGTGCC...TGCATAACCCCTCATCTCTCGGAGGCG 216
| ||| :|||:|||||:|||||:|||||:|||||:|||||
50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66
|:|:| | ||| :|||:|||||:|||||:|||||:|||||: |||
217 TAGCTCTCCAGCAATATCGAGGAGGGCTAACCTCACCACACTCCCGGA 266
| ||| :|||:|||||:|||||:|||||:|||||:|||||
66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83
| ||| :|||:|||||:|||||:|||||:|||||:|||||
267 GAACGGCACAATTGTGGTGAACATGCCCAGCTGAGCCAGGATGACTCG 316
| ||| :|||:|||||:|||||:|||||:|||||:|||||
83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
|:|:| | ||| :|||:|||||:|||||:|||||:|||||: |||
317 GCGCGTACAAGTGTGCGCTGGGCATCAATAGC...CGAGGCGCTGCTT 363
| ||| :|||:|||||:|||||:|||||:|||||:|||||
100 LysValThrLeuAsnVal 105
| ||| :|||:|||||:|||||:|||||:|||||:|||||
364 GATGTCAGCTGGAGGTC 381

seq_name: N_Geneseq_36:T30857

seq_documentation_block:

ID T30857 standard; cDNA; 1936 BP.
AC T30857;
DT 13-SEP-1996 (first entry)
DE Secretary component DNA fragment 2.
KW Polymeric immunoglobulin receptor; pIgR; CHO; ss.
OS Chimeric Homo sapiens;
OS Chimeric synthetic.
FH Key Location/Qualifiers
FT misc_feature 1..46
FT /tag= a
FT /note= "bases 1-46 derived from PCR with 5'
FT primer 2"
FT 16..1920
FT /tag= b
FT mat_peptide
FT /product= transcript from SC DNA fragment 2
FT complement (1902..1936)
FT /tag= c
FT /note= "from PCR with 3' primer"
PN WO9618734-A1.
PD 20-JUN-1996.
PE 06-DEC-1995; E04797.
PR 16-DEC-1994; EP-120019.
PA (CIBA) CIBA GEIGY AG.
PI Asselbergs FAM, Gruetter M, Leist C, Van Oostrum J;
DR WPI; 96-300651/30.
PT Prodn. of recombinant secretory protein capable of crystallisation -
PT pref. human poly-IgR, for use in screening and binding studies
PS Claim 5; Page 39-41; 50pp; English.
CC SC DNA fragment 2 (T30857) was obt'd. by PCR amplification of human
CC polymeric immunoglobulin receptor (pIgR) cDNA in vector pCB6 using a
CC 5' primer (T30860) that creates an HindIII site upstream of the ATG
CC start codon and a 3' primer (T30861) that generates a stop codon at
CC position 1906 just upstream of the hydrophobic transmembrane segment
CC of the pIgR and also creates a downstream XbaI site. CHO SSF 3
CC cells transfected with a vector carrying the amplified DNA can be
CC used for prodn. of recombinant secretory component (SC). This can
CC be crystallised for receptor structure studies, used to stabilise
CC immunoglobulins, or used to screen (ant)agonists capable of
CC modulating mucosal immune responses.
SQ Sequence 1936 BP; 452 A; 525 C; 578 G; 381 T;

alignment_scores:

Quality: 180.50 Length: 106
Ratio: 2.314 Gaps: 4
Percent Similarity: 73.585 Percent Identity: 38.679

alignment_block:

US-09-135-238B-2_COPY_18_253 x T30857 ..

Align seg 1/1 to: T30857 from: 1 to: 1936

4 ProGluValLysValGluGlyGluLeuGlySerValThrIleLysCys 20
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
88 CCGGAG...GAGGTGAATAGTGTGAAGGTAACCTCAGTGTCCATCAGCTG 134
| ||| :|||:|||||:|||||:|||||:|||||:|||||
20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33
| ||| :|||:|||||:|||||:|||||:|||||:|||||
135 CTACTACCCACCCACCTCTGTCAACCGGCACACCGGAAGTACTGGTGCC 184
| ||| :|||:|||||:|||||:|||||:|||||:|||||
33 rGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
|:|:| | ||| :|||:|||||:|||||:|||||:|||||: |||
185 GGCAGGGAGCTAGAGTGCC...TGCATAACCCCTCATCTCTCGGAGGCG 231
| ||| :|||:|||||:|||||:|||||:|||||:|||||
50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66
|:|:| | ||| :|||:|||||:|||||:|||||:|||||: |||
232 TAGCTCTCCAGCAATATCGAGGAGGGCTAACCTCACCACACTCCCGGA 281
| ||| :|||:|||||:|||||:|||||:|||||:|||||
66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83
| ||| :|||:|||||:|||||:|||||:|||||:|||||

282 GAACGGCACATCTCGTGAACATTCGCCACCTGAGCGAGGATGACTCCG 331
 83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
 || ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 332 GCGCTACAAGTGTGGCTGGCATCAATAGC...CGAGGCGTGTCTTT 378
 100 LysValThrLeuAsnVal 105
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 379 GATGTCAGCCTGGAGGTC 396

seq_name: N_Geneseq_36:T30856

seq_documentation_block:
 ID T30856 standard; cDNA; 2031 BP.
 AC T30856;
 DT 13-SEP-1996 (first entry)
 DE Secretory component DNA fragment 1.
 KW Secretory component; SC DNA fragment 1; crystallisation;
 KW polymeric immunoglobulin receptor; pIgR; CHO; ss.
 OS Chimeric Homo sapiens;
 OS Chimeric synthetic.
 FH Key Location/Qualifiers
 FT misc_feature 1..33
 FT /tag= a
 FT /note= "bases 1-33 derived from PCR with 5'
 primer 1"
 FT primer 1
 FT 111..2015
 FT /tag= b
 FT /product= transcript from SC DNA fragment 1
 FT complement (1997..2031)
 FT /tag= c
 FT /note= "from PCR with 3' primer"
 FT WO9618734-A1.
 PN 20-JUN-1996.
 PD 06-DEC-1995; E04797.
 PR 16-DEC-1994; EP-120019.
 PA (CIBA) CIBA GEIGY AG.
 PI Asselbergs FM, Gruetter M, Leist C, Van Oostrum J;
 DR WPI; 96-300651/30.
 DT Prod. of recombinant secretory protein capable of crystallisation -
 pref. human poly-IgR, for use in screening and binding studies
 PT Claim 5; Page 36-38; 50pp; English.
 PS SC DNA fragment 1 (T30856) was obt'd. by PCR amplification of human
 CC polymeric immunoglobulin receptor (pIgR) cDNA in vector pCB6 using a
 CC 5' primer (T30859) hybridising to the cytomegalovirus promoter in
 CC pCB6 and a 3' primer (T30861) that generates a stop codon at
 CC position 1906 just upstream of the hydrophobic transmembrane segment
 CC of the pIgR and also creates a downstream XbaI site. CHO SSF 3
 CC cells transfected with a vector carrying the amplified DNA can be
 CC used for prodn. of recombinant secretory component (SC). This can
 CC be crystallised for receptor structure studies, used to stabilise
 CC immunoglobulins, or used to screen (ant)agonists capable of
 CC modulating mucosal immune responses.
 SQ Sequence 2031 BP; 477 A; 541 C; 605 G; 408 T;

alignment_scores:
 Quality: 180.50 Length: 106
 Ratio: 2.314 Gaps: 4
 Percent Similarity: 73.585 Percent Identity: 38.679

alignment_block:

US-09-135-238B-2_COPY_18_253 x T30856 ..

Align seq 1/1 to: T30856 from: 1 to: 2031

4 ProGluValLysValGluGlyGluLeuValThrIleIleYscY 20
 183 CCGAG...GAGGTGAATAGTGTGGAGGTAACTGATGTCATCACTGCTG 229
 20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33
 || ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 230 CTACTACCCACCCACTCTGTCAACCGGCACACCCCGAAGTACTGTGTC 279

33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
 ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 280 GCGAGGAGGATGAGGTGGC...TGCATAACCCCTCATCTCTCGGAGGCG 326
 50 PheIleLysAlaGluTyrLysGlyAlaGlyValThrLeuLysGlnTyrProAr 66
 :|||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 327 TACGTCACCAATATGACGAGGCGGCTAACCTCACCACACTCCCGGA 376
 66 GlyAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83
 :|||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 377 GAACGGCACATTCGTGTGAACATTGCCAGCTGAGCCAGCATGACTCCG 426
 83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
 || ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 427 GCGCTACAAGTGTGGCTGGCATCANTAGC...CGAGGCGTGTCTTT 473
 100 LysValThrLeuAsnVal 105
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 474 GATGTCAGCCTGGAGGTC 491

seq_name: N_Geneseq_36:T31288

seq_documentation_block:

ID T31288 standard; cDNA; 2919 BP.
 AC T31288;
 DT 24-FEB-1997 (first entry)
 DE Human poly-immunoglobulin receptor, cDNA.
 KW Human; immunoglobulin receptor; protection protein; mutants;
 KW heavy chain; antigen binding domain; protection; pathogen;
 KW mucosal; environment; gastrointestinal; passive; immunisation;
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
 KW poly; sorbinus; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 235..2475
 FT /*tag= a
 FT WO9621012-A1.
 PN 11-JUL-1996.
 PD 27-DEC-1995; U16889.
 PR 30-DEC-1994; US-367395.
 PR 04-MAY-1995; US-434000.
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.
 PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
 PA (PLAN-) PLANET BIOTECHNOLOGY INC.
 PI Hiatt AC, Leher T, Ma JKC;
 DR WPI; 96-333987/33.
 DR P-PSDB; W03178.
 PT Immunoglobulin and protection protein complex and its prodn. in
 PT plants - useful for passive immunisation against mucosal antigens,
 PT esp. against S. mutants and S. sorbinus to prevent dental caries
 PS Disclosure: Pages 105-108; 152pp; English.
 CC The present sequence encodes the human poly-immunoglobulin (Ig)
 CC receptor, a portion of which corresp. to residues 1-627, pref.
 CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
 CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
 CC The Ig of the invention comprises a PP as above in association with
 CC an Ig derived heavy chain, having at least a portion of an antigen
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
 CC gastrointestinal, environments, therefore enhancing its
 CC effectiveness in passively immunising animals against mucosal
 CC pathogens. The Ag binding domain is specifically derived from the
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
 CC by binding, e.g. Streptococcus mutants serotypes c, e and f, or
 CC S. sorbinus serotypes d and g.
 SQ Sequence 2919 BP; 684 A; 806 C; 854 G; 575 T;

alignment_scores:
 Quality: 180.50 Length: 106
 Ratio: 2.314 Gaps: 4
 Percent Similarity: 73.585 Percent Identity: 38.679

alignment_block:

```
US-09-135-238B-2_COPY_18_253 x T31288 ..
Align seg 1/1 to: T31288 from: 1 to: 2919

4 ProGluValLysValGluGlyLeuGlySerValThrIleLysCys 20
253 CCGGAG...GAGGTGAATATGTGGAAGTAACTCAGTGTCCATCACGTG 299
||||| :|||:||||| :|||:|||||:|||||:|||||:
20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33
||||| :|||:||||| :|||:|||||:|||||:|||||:
300 CTACTACCCACCACCTCTGTCAACGGCACCCGGAAGTACTGGTGCC 349
||||| :|||:||||| :|||:|||||:|||||:|||||:
33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
||||| :|||:||||| :|||:|||||:|||||:|||||:
350 GGCAGGGAGCTAGAGTGCC...TGCATAACCTCTATCTCTCGGAGGCG 396
||||| :|||:||||| :|||:|||||:|||||:|||||:
50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66
||||| :|||:||||| :|||:|||||:|||||:|||||:
397 TAGCTCTCCAGCAATATGCGAGGAGGCTAACCTCACCACTTCCCGGA 446
||||| :|||:||||| :|||:|||||:|||||:|||||:
66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerg 83
||||| :|||:||||| :|||:|||||:|||||:|||||:
447 GAACGGCACATTGTGGTGAACATTGCCAGCTGAGCCAGGATGACTCCG 496
||||| :|||:||||| :|||:|||||:|||||:|||||:
83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
||||| :|||:||||| :|||:|||||:|||||:|||||:
497 GCGCGTACAAAGTGTGGCTGGGCATCAATAGC...CGAGGCGTGTCTTT 543
||||| :|||:||||| :|||:|||||:|||||:|||||:
100 LysValThrLeuAsnVal 105
||||| :|||:||||| :|||:|||||:|||||:|||||:
544 GATGTCAGCGCTGGAGGTC 561
||||| :|||:||||| :|||:|||||:|||||:|||||:
```

seq_name: N_Geneseq_36:T31289

seq_documentation_block:

```
ID T31289 standard; cdNA; 3630 BP.
AC T31289;
DT 24-FEB-1997 (first entry)
DE Bovine poly-immunoglobulin receptor, cdNA.
KW Bovine; Immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; cow; ss.
OS Bos taurus.
FH Key Location/Qualifiers
FT cds 152..2425
FT /*tag= a
PN WO9621012-A1.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI; 96-333987/33.
DR P-PSDB; W03179.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure; Pages 111-115; 152pp; English.
CC The present sequence encodes the bovine poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
```

```
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 3630 BP; 880 A; 1018 C; 1029 G; 703 T;
```

alignment_scores:

```
Quality: 179.50 Length: 109
Ratio: 2.272 Gaps: 4
Percent Similarity: 72.477 Percent Identity: 37.615
```

alignment_block:

US-09-135-238B-2_COPY_18_253 x T31289 ..

Align seg 1/1 to: T31289 from: 1 to: 3630

```
4 ProGluValLysValGluGlyGluLeuGlyGlySerValThrIleLysCys 20
||||| :|||:||||| :|||:|||||:|||||:|||||:
224 CCGGAG...GAGGTGAGCAGCGTGGAAAGCGCGCTCAGTGTCCATCAAGTG 270
||||| :|||:||||| :|||:|||||:|||||:|||||:
20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33
||||| :|||:||||| :|||:|||||:|||||:|||||:
271 CTACTACCCGCCACCTCCGTCAACGGGCACACGCGCAAGTACTGGTGCC 320
||||| :|||:||||| :|||:|||||:|||||:|||||:
33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
||||| :|||:||||| :|||:|||||:|||||:|||||:
321 GGCAGGAGGCC...CAGGGCGCGCTGCAGCACCTCATCTCTCGGAGGCG 367
||||| :|||:||||| :|||:|||||:|||||:|||||:
50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66
||||| :|||:||||| :|||:|||||:|||||:|||||:
368 TAGCTCTCCGACGACTACGTGGCGAGAGCAACCTCACCAACTTCCCGGA 417
||||| :|||:||||| :|||:|||||:|||||:|||||:
66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerg 83
||||| :|||:||||| :|||:|||||:|||||:|||||:
418 GAGCGGCACGTTTGTGTGGTGACATCAGCCATCTCACCCATAAAGACTCAG 467
||||| :|||:||||| :|||:|||||:|||||:|||||:
83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
||||| :|||:||||| :|||:|||||:|||||:|||||:
468 GGCCTCAACAAGTGTGGCTGGGCATTAGCAGC...CGTGGCCTTAACCTTC 514
||||| :|||:||||| :|||:|||||:|||||:|||||:
100 LysValThrLeuAsnValHisSergGlu 108
||||| :|||:||||| :|||:|||||:|||||:|||||:
515 GATGTGAGCCTGGAGGTCAGCCCAAGAT 541
||||| :|||:||||| :|||:|||||:|||||:|||||:
```

seq_name: N_Geneseq_36:V20383

seq_documentation_block:

```
ID V20383 standard; cdNA to mRNA; 1911 BP.
AC V20383;
DT 26-JUN-1998 (first entry)
DE cdNA for human immunity related factor.
KW Lymph node; human; immunity related factor; research; treatment;
KW Immune disease; infectious disease; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 98..1696
FT /*tag= a
FT sig_peptide 98..145
FT mat_peptide 146..1693
FT /*tag= b
FT /*tag= c
FT /product= immunity_related_factor
PN J10072495-A.
PD 17-MAR-1998.
PF 11-JUN-1997; 153218.
PR 13-JUN-1996; JP-152362.
PA (ASAH ) ASAH KASEI KOGYO KK.
DR WPI; 98-234766/21.
DR P-PSDB; W50033.
PT Immunity related factor - useful in the treatment of immune related
PT and infectious diseases
PS Claim 15; Pages 18-20; 21pp; Japanese.
```

CC The present sequence encodes a lymph node derived human immunity
CC related factor, which can be used to research and treat immune and
CC infectious diseases.
SQ Sequence 1911 BP; 490 A; 541 C; 525 G; 355 T;

alignment_scores:
Quality: 175.50 Length: 306
Ratio: 1.415 Gaps: 10
Percent Similarity: 40.523 Percent Identity: 21.895

alignment_block:

US-09-135-238B-2_COPY_18_253 x V20383 ..

Align seg 1/1 to: V20383 from: 1 to: 1911

```
8 valGluGlyGluLeuLeuGlySerValThrIleLysCysProLeuPro.. 23
|||||
305 GTGTCAGGGAGCGCTGGAGGAGCTGTCCACCATCGTGCCTATTATGCCCC 354
24 .....GluMethHisValArgIleTyrLeuCysArgGluMetAlaG 37
   :   :   :   :   :   :   :   :   :   :   :   :   :
355 CTCATCTGTCAACAGGCACGACGAGTACTGTCGCTGTGGGCCCC 404
37 lysSerGlyThrCysGlyThrValSerThrThrAsnPheIleLysAla 53
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
405 CAAGATGGATCTGCCAGACCATGTGTCCCAACACCATATATCACCAT 454
54 GluTyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPh 70
   :   :   :   :   :   :   :   :   :   :   :   :   :
455 CGCTATCTGACCGTGTGGCCCTACAGACTTCCACAGAGAGCTGT 504
70 eLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaC 87
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
505 TGTGTGAGGCTGTCCCAACTGTCCCGGATGACATCGATCTACCTCT 554
87 ysGlyAlaGlyMetAsnThrAspArg..... 95
555 GCGGATTGGAGTGAACACACATGCTGTCTTAAGCATGAATGACGCC 604
95 ..... 95
605 ATCTCTCAGGTCCCGCCAGCACCTCCACACGCCACTCCAGCTGTGG 654
95 ..... 95
655 GGAGCTCACCATGAGATCTTATGGACAGCGCTCCAGTGGCCACAGAT 704
96 .....GlyLysThrGln..... 99
705 GGACCCCGAGGAACCCAGACCTTAGCAGGGGACAGCATGGGACACA 754
100 .....LysValThrLeuAsnValHisSerG1 108
755 GTTGCTTCCATCCAGACCCAGCAGACTACACTTACGCTGAGGGAAG 804
108 uTyrGluPro.....SerTrp 114
805 ACGAACCCCGAGGACACCCAGGCGCAGCTCCAGGGACAGCGCTGGG 854
114 LuGlu.....GlnProMetProGluThrProLysTrp 124
855 CAGAGGGTTCGTCAAGACACCTGCTCCGATTCAGAGAGATCCA..... 898
125 PheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerLys 141
899 .....CCTTCAAGAGCAG 912
141 sPheValThrArgValThrThrProAlaGlnArgGlyLysValProProv 158
913 AAGCATGTCCATACACAGAGAGTGTGGAGGGC..... 949
158 aHisHisSerSerProThrThrGlnIleThrHisArgProArgValSer 174
```

```
950 .....ACCAAGCTCGGTGACAAACGGGCTAGAGCCAC 985
175 Arg .....AlaSerSerValAlaGlyAspLysProArg..... 185
986 AAGGACAGGAGGAGATGACAACCTACCAAGGCTGATAGGCCAAGGAGGA 1035
186 .....T 186
1036 CATAGAGGGGTCAGGATAGCTTTGATGACCCAAAAAGGCTCTAGGAA 1085
186 hrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuGluGlyLeu 202
1086 CCATTGGGCCACGAGCTCTGTCTCTCAGAAACTTTGGCTGGGAA...ATC 1132
203 LeuLysProGlnThrPro 208
1133 CTCCCAACAGCAACGCCA 1150
seq_name: N_Geneseq_36:Q14498
```

seq_documentation_block:

ID Q14498 standard; DNA; 1876 BP.

AC Q14498;

DT 10-MAR-1993 (revised)

DE 30-JAN-1992 (first entry)

KW Rabbit; insemination; pregnancy; ss.

OS Oryctolagus cuniculus.

FT Key Location/Qualifiers

PN W09116061-A.

PD 31-OCT-1991.

PF 16-APR-1991; U02604.

PR 16-APR-1990; US-510161.

PA (HARD) HARVARD COLLEGE.

PI (SURE-) INST SUISSE RECH EXPER C.

DR WPI; 91-339549/46.

PT P-PSDB; RI4670.

PR Stabilised poly-Ig complex contg. portion of poly-Ig receptor -

PS Disclosure; Fig 8; 51 pp; English.

CC The sequence was obt'd. from a cDNA clone of allele no. 1 and

CC encodes a truncated poly-Ig receptor. The native gene (Mostov et

CC al) is mutated to delete the portion encoding the transmembrane and

CC intra-cellular domains. The recombinant protein produced by

CC expression of the sequence is used as a stabiliser protein with a

CC poly-Ig specific for a selected antigen or family of antigens. The

CC compn. can be administered directly to the mucosal surfaces of a

CC mammal to protect against a pathogen or against insemination. It

CC protects against allergens that contact the respiratory or digestive

CC mucosal surfaces and protects against pregnancy by cross-linking

CC sperm in the vagina.

CC See also Q14499.

SQ Sequence 1876 BP; 441 A; 525 C; 549 G; 361 T;

alignment_scores:

Quality: 166.50 Length: 103

Ratio: 2.602 Gaps: 4

Percent Similarity: 62.136 Percent Identity: 42.718

alignment_block:

US-09-135-238B-2_COPY_18_253 x Q14498 ..

Align seg 1/1 to: Q14498 from: 1 to: 1876

13 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 25

238 GCGCACTCGGTGTCATCATGCTACTACCCCAACCTCCGTCACCGC 287

25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42


```

288 GCACAGCGGAGTCTGTCGCGCGGAA...GAGGAGAGCGCGCGTCGC 334
42 lyThrValValSerThrThrAsnPhelLysAlaGluTyrLysGlyArg 58
335 TGACGCTGCTCG...ACGGCTACAGCTCCAGGAATACTCCGGGAGA 381
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
382 GGCAAGCTCACCGACTTCCCTGATAAGGGAGTTTGTGTGACTGTGA 431
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
432 CCAACTCACCCAGAGCACTCAGGAGCTACAAGTGTGGTGGGAGTCA 481
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
482 AC...GGCCGTGGCTGGACTTCGGTGTCAACGTGCTGAGCCAGGAAG 528
109 TyrGluPro 111
529 CCAGAGCCT 537

```

seq_name: N_Geneseq_36:T31287

seq_documentation_block:
ID T31287 standard; cdna; 3517 BP.

```

AC T31287;
DT 24-FEB-1997 (first entry)
DE Rabbit poly-immunoglobulin receptor, cDNA.
KW Rabbit; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; ss.
OS Oryctolagus cuniculus.
FH Key Location/Qualifiers
FT cds 124..2445
FT /tag= a
PD WO9621012-Al.
PN 11-JUL-1996.
PR 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNWE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI; 96-333987/33.
DR P-PSDB; W03177.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutants and S. sorbinus to prevent dental caries
PS Claim 10; Pages 99-102; 152pp; English.
CC The present sequence encodes the rabbit poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 21-43, 1-118, 119-223, 224-332, 333-441,
CC 442-552, 553-606 or 553-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutants serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 3517 BP; 795 A; 1027 C; 1038 G; 657 T;

```

alignment_scores:
Quality: 166.50 Length: 103
Ratio: 2.602 Gaps: 4

Percent Similarity: 62.136 Percent Identity: 42.718

alignment_block:

US-09-135-238B-2_COPY_18_253 x T31287 ..

Align seg 1/1 to: T31287 from: 1 to: 3517

```

13 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 25
111 |||||:||||| |||
238 GCGACTCGGTGTCCATCATCTACTACCAACAACCTCCGTCACCG 287
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
111 |||: |||: ||||| |||
288 GCACAGCGGAGTTCGTGCGCGGAA...GAGGAGAGCGCGCGTCGC 334
42 lyThrValValSerThrAsnPhelLysAlaGluTyrLysGlyArg 58
111:|||||: |||: |||: |||: |||: |||: |||: |||
335 TGACGCTGCTCG...ACGGCTACAGCTCCAGGAATACTCCGGGAGA 381
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
111 |||: |||: |||: |||: |||: |||: |||
382 GGCAAGCTCACCGACTTCCCTGATAAGGGAGTTTGTGTGACTGTGA 431
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
111:|||||: |||: |||: |||: |||: |||: |||: |||
432 CCAACTCACCCAGAGCACTCAGGAGCTACAAGTGTGGTGGGAGTCA 481
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
11 |||: |||: |||: |||: |||: |||: |||: |||
482 AC...GGCCGTGGCTGGACTTCGGTGTCAACGTGCTGAGCCAGGAAG 528
109 TyrGluPro 111
529 CCAGAGCCT 537

```

seq_name: N_Geneseq_36:Q14499

seq_documentation_block:

```

ID Q14499 standard; cdna; 1876 BP.
AC Q14499;
DT 30-JAN-1992 (first entry)
DE Clone 2 for truncated poly Ig-receptor.
KW Rabbit; insemination; pregnancy; ss.
OS Oryctolagus cuniculus.
FH Key Location/Qualifiers
FT cds 124..1876
PN W09116061-A.
PD 31-OCT-1991.
PR 16-APR-1991; U02604.
PR 16-APR-1990; US-510161.
PA (HARD ) HARVARD COLLEGE.
PA (SURE-) INST SUISSE RECH EXPER C.
PI Kraehenbuhl JP, Weltzin RA, Neutra MR;
DR WPI; 91-339549/46.
DR P-PSDB; R14671.
PT Stabilised poly-Ig complex contg. portion of poly-Ig receptor -
PT useful in protection against pathogens or against pregnancy
PT Disclosure; Fig 8; 51 pp; English.
CC The sequence was obt'd. from a cDNA clone of allele no. 2 and
CC encodes a truncated poly-Ig receptor. The native gene (Mostov et
CC al) is mutated to delete the portion encoding the transmembrane and
CC intra-cellular domains. The recombinant protein produced by
CC expression of the sequence is used as a stabiliser protein with a
CC poly-Ig specific for a selected antigen or family of antigens. The
CC compsn. can be administered directly to the mucosal surfaces of a
CC mammal to protect against a pathogen or against insemination. It
CC protects against allergens that contact the respiratory or digestive
CC mucosal surfaces and protects against pregnancy by cross-linking
CC sperm in the vagina.
CC See also Q14498.
SQ Sequence 1876 BP; 357 A; 452 C; 483 G; 306 T;

```

CC produced.. The same method can be used to identify antagonists of PIGR-2
CC but the presence of an antagonist is indicated by a decrease in
CC production of the signal. Antibodies against PIGR-2 may be used to
CC isolate or identify clones expressing PIGR-2.
SQ Sequence 1137 BP; 252 A; 318 C; 326 G; 241 T;

alignment_scores:
Quality: 119.50 Length: 153
Ratio: 1.532 Gaps: 7
Percent Similarity: 50.980 Percent Identity: 27.451

alignment_block:
US-09-135-238B-2_COPY_18_253 x X28250 ..

Align seg 1/1 to: X28250 from: 1 to: 1137

Align seg 1/1 to: Q14499 from: 1 to: 1876

8 ValGluGlyGluLeuGlyGlySerValThrIleLysCysProLeuProG1 24
216 GTGACTGGCAGTGGCGGGGACTCTCTGACAGTGGTGTGTATGAGAG 265
24 uMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerG 39
266 CATGTACAGGGATATACAGTACTGTGGTCCGAGGACAGTACGACAG. 314
39 lyThrCysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyr 55
315 ..TCATGTGAGAGCATTTGGAGACCAAGGAGGAGAGAGAGTGGAGAG 362
56 LysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuVa 72
363 AATGGCGCGTGTCCATCAGAGACCCCGGAGGCTCTCGCCTTACTGT 412
72 IgluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyA 89
413 GACCATGAGAACCTCAATGAAGATGATGCTGGATCTTACTGTGC.... 458
89 laGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 105
459AAAAATTCAGACAGTGTGGTGC..... 479
106 HisSerGluTyrGluProSerTrpGluGluGlnProMetProGluThrPr 122
480CTGGATTTCATGTCACGGATCCCTCG..... 506
122 olyStrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerS 139
507GACCTGGTTAGGGTGTATGTTTCCCGACCAATTACACCC 546
139 erSerLysPheValThrArgValThrThrProAlaGlnArgGlyLysVal 155
547 CA.....AGGAGGACACACATCCAGCC.....ACA 572
156 ProProVal 158
573 CCTCCCATC 581

13 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 25
238 GGGACTGGTGTCCATCAGTGTACTACCAACACCTCGGTCAACCG 287
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
288 GCACAGCCGGGAAGTTCTGTGGTCCGGGAA...GAGGAGAGCGCGCTGCG 334
42 lyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 58
335 TGACGCTTGCTCG...ACGGCTACAGTCCCGAGGAACTACTCCGGGAGA 381
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
382 GGCATCTCAGCGACTTCCCTGAAAGGGGAGTTTGTGTGTATTATAA 431
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
432 CCACACTCAGCAGAGCGACTCAGGGAGTACAGTGTGGGTGGAGTCA 481
92 sn 92
482 AC 483

seq_name: N_Geneseq_36:X28250

seq_documentation_block:
ID X28250 standard; cDNA; 1137 BP.
AC X28250;

DE 16-JUN-1999 (first entry)
DE Human PIGR-2 coding sequence.
KW PIGR-2; human; autoimmune disease; rheumatoid arthritis; psoriasis;
KW Multiple Sclerosis; Systemic Lupus Erythematosus; diagnosis; therapy;
KW Inflammatory Bowel Disease; SS.
OS Homo sapiens.
PN EP-905237-A2.
PD 31-MAR-1999.
PR 21-AUG-1998; 306323.
PR 07-AUG-1997; US-976293.
PR 25-AUG-1997; US-056774.
PA (SMIK) SMITHKLINE BEECHAM CORP.
FI Sweet RW, Truneh A, Wu S;
DR WPI; 99-192665/17.
DR P-PSDB; Y05069.
DR New polypeptides encoding human PIGR-2 useful for treating diseases
PT such as rheumatoid arthritis and multiple sclerosis
PS Claim 4; Page 16-17; 23pp; English.
CC This sequence encodes the human PIGR-2 protein of the invention.
CC Autoimmune diseases involving altered expression or activity of PIGR-2
CC may include rheumatoid arthritis, Multiple Sclerosis, psoriasis,
CC Systemic Lupus Erythematosus and Inflammatory Bowel Disease. These
CC diseases can be diagnosed or susceptibility to them predicted by:
CC (1) determining whether there is a mutation in the genomic copy of the
CC gene encoding PIGR-2; or (2) measuring the amount of PIGR-2 in a sample
CC derived from the patient. Patients deficient in PIGR-2 can be treated by
CC administering either the PIGR-2 DNA or its complement or an agonist of
CC PIGR-2 to the patient. Patients with excessive expression or activity of
CC PIGR-2 can be treated by administering an antagonist of PIGR-2, an
CC antisense nucleic acid molecule which inhibits the expression of PIGR-2
CC or administering sufficient PIGR-2 to compete with the endogenous
CC activity. PIGR-2 can be used to identify its agonists by contacting a
CC cell expressing PIGR-2 with a candidate compound in the presence of a
CC signal system and noting the candidate as an agonist if a signal is

